

UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS

FACULTAD DE CIENCIAS MATEMÁTICAS

E.A.P. DE ESTADÍSTICA

**Modelos espaciales lineales gaussianos en el estudio de la
variabilidad espacial**

TESIS

Para optar el Título Profesional de Licenciado en Estadística

AUTOR

Miguel Angel Uribe Opazo

ASESORA

Emma Cambillo Moyano

Lima – Perú

2015

FICHA CATALOGRAFICA

MIGUEL ANGEL URIBE OPAZO

**Modelos Espaciales Lineales Gaussianos en el Estudio de la Variabilidad Espacial.
(Lima, 2015).**

Xxx, 80p.xxxx, UNMSM, Licenciado, Estadística, 2015.

Tesis Universidad Nacional Mayor de San Marcos

Facultad de Ciencias Matemáticas 1. Estadística

**I. UNMSM/FdeCMII. Modelos Espaciales Lineales Gaussianos en el Estudio de la
Variabilidad Espacial.**

MODELOS ESPACIALES LINEALES GAUSIANOS EN EL ESTUDIO DE LA
VARIABILIDAD ESPACIAL

MIGUEL ANGEL URIBE OPAZO

Tesis presentada a consideración del Cuerpo Docente de la Facultad de Ciencias Matemáticas de la Universidad Nacional Mayor de San Marcos, como parte de los requisitos para obtener el Título Profesional de Licenciado en Estadística.

Aprobada por:

Mg. Ysela Agüero Palacios

Mg. Maria Estela Ponce Aruneri

Mg. Emma Norma Cambillo Moyano

Lima-Perú

Mayo-2015

AGRADECIMIENTOS

Agradezco a Dios, por todas las cosas de esta vida que me continua dando con mucho amor.

A mi orientadora la Profesora Emma Cambillo Moyano por toda confianza y constante apoyo para terminar este trabajo.

A todos mis profesores de la Escuela Académica Profesional de Estadística de la Universidad Nacional Mayor de San Marcos que contribuyeron con mi formación.

A la Profesora Edith Seier por todo el incentivo y apoyo en todos estos años.

A toda mi familia, que siempre me incentivo a continuar y a concluir esta etapa de mi vida.

A mis amigos y colegas por el incentivo y amistad.

A todos, muchas gracias.

MODELOS ESPACIALES LINEALES GAUSIANOS EN EL ESTUDIO DE LA VARIABILIDAD ESPACIAL

MIGUEL ANGEL URIBE OPAZO

Mayo-2015

Orientadora: Mg. Emma Cambillo Moyano
Título Obtenido: Licenciado en Estadística

RESUMEN

Los modelos espaciales lineales gaussianos son utilizados para explicar la estructura de variabilidad espacial de variables aleatorias regionalizadas, permitiendo la construcción de mapas temáticos por interpolación utilizando la técnica kriging. La estimación de los parámetros del modelo es realizada utilizando el método de máxima verosimilitud, obteniendo estimadores no sesgados de variancia mínima. Sin embargo, la estimación de los parámetros del modelo, así como la construcción de los mapas temáticos, puede ser afectada por la presencia de observaciones influyentes, siendo necesario estudios de sensibilidad. El objetivo de este trabajo es de presentar: los métodos geoestadísticos en el estudio de la variabilidad espacial de variables regionalizadas utilizando los modelos espaciales lineales gaussianos; estudiar las técnicas de diagnósticos de influencia global y local de las variables regionalizadas utilizando tres esquemas de perturbación para la variable respuesta que son presentadas en la literatura. Finalmente, la metodología presentada es aplicada en dos conjuntos de datos reales, posibilitando evaluar la metodología desarrollada.

Palabras clave: Dependencia espacial, Diagnóstico, Geoestadística, Máxima verosimilitud.

ABSTRACT

GAUSSIAN SPATIAL LINEAR MODELS IN THE STUDY OF SPATIAL VARIABILITY

AUTHOR: Bach. MIGUEL ANGEL URIBE OPAZO

ADVISED: Mg. EMMA NORMA CAMBILLO MOYANO

Gaussian spatial linear models are used to explain the structure of spatial variability of regionalized random variables. Allowing the construction of thematic maps by interpolation using the technique known as kriging. The estimation of the model parameters is performed using the maximum likelihood method estimators, obtaining estimators non biased and with minimum variance. Nevertheless, the estimation of the model parameters as well as the construction of the thematic maps can be affected by the presence of influential observations, so diagnostics techniques are needed. The goal of this work is to present: the geostatistical method in the study of the spatial variability of regionalized variables using Gaussian linear spatial models; diagnostics techniques to study global and local influence using three perturbation schemes for the response variable which are presented in the literature. Finally, the methodology presented is applied on two real datasets, allowing to evaluate the methodology developed.

Keywords: Diagnostics, Geostatistics, Maximum likelihood, Spatial dependence.

INDICE

| | |
|---------------------------------------------------------------------------------|------------|
| LISTA DE TABLAS | III |
| LISTA DE FIGURAS | IV |
| INTRODUCCIÓN | 1 |
| 1. GEOESTADÍSTICA..... | 5 |
| 1.1 Teoría de las variables regionalizadas | 6 |
| 1.2 Semivariograma | 8 |
| 1.3 Anisotropía | 12 |
| 1.4 Modelos espaciales..... | 13 |
| 1.5 Estimación de parámetros en el ajuste de modelos teóricos a semivariogramas | 17 |
| 1.5.1 Método de mínimos cuadrados ordinarios (OLS) | 17 |
| 1.5.2 Método de mínimos cuadrados generalizados (GLS) | 18 |
| 1.5.3 Método mínimos cuadrados ponderados (WLS)..... | 18 |
| 1.5.4 Método de máxima verosimilitud (ML) | 21 |
| 1.5.5 Método de máxima verosimilitud restringida (MLR) | 21 |
| 1.6 Kriging..... | 22 |
| 1.6.1 Kriging ordinario | 23 |
| 1.6.2 Kriging universal..... | 26 |
| 1.7 Criterios de selección de modelos..... | 27 |
| 1.7.1 Validación cruzada | 27 |
| 1.7.2 Criterios de información de Akaike (<i>AIC</i>) | 29 |
| 1.7.3 Criterios de información de Schwarz (<i>SIC</i>)..... | 30 |
| 2. COMPARACIÓN DE MAPAS TEMÁTICOS..... | 31 |
| 3. MODELO ESPACIAL LINEAR GAUSIANO..... | 34 |
| 3.1 Función score | 35 |
| 3.2 Matriz de información observada | 36 |
| 3.3 Matriz de información esperada | 36 |
| 4. DIAGNÓSTICO | 38 |
| 4.1 Influencia global | 38 |

| | | |
|-----------|-------------------------------------------------------------------------------------|-----------|
| 4.2 | Influencia local | 39 |
| 4.3 | Influencia local conformal | 43 |
| 5. | APLICACIONES | 44 |
| 5.1 | Caracterización del área I en estudio | 44 |
| 5.2 | Caracterización del área II en estudio | 45 |
| 5.3 | Análisis de los datos | 46 |
| 5.4 | <i>Software</i> utilizado | 47 |
| 6. | RESULTADOS Y DISCUSIONES | 48 |
| 6.1 | Análisis del área I: Estudio de la resistencia del suelo a la penetración - RSP ... | 48 |
| 6.1.1 | Análisis exploratorios de datos | 48 |
| 6.1.2 | Análise de la estructura de dependencia espacial y criterios de validación | 50 |
| 6.1.3 | Construcción de mapas temáticos..... | 52 |
| 6.1.4 | Dagnósticos de Influencia local | 52 |
| 6.1.5 | Análisis de la RSP sin considerar la observación influyente | 54 |
| 6.2 | Análisis del Área II: Estudio de la Productividad de trigo | 58 |
| 6.2.1 | Modelo espacial lineal | 58 |
| 6.2.2 | Análisis exploratorios de los datos | 59 |
| 6.2.3 | Análisis espacial..... | 60 |
| 6.2.4 | Diagnósticos de Influencia global y local | 63 |
| 6.2.5 | Comparación de mapas temáticos | 66 |
| 7. | CONCLUSIONES | 70 |
| | BIBLIOGRAFIA | 71 |
| | ANEXO A..... | 77 |
| | ANEXO B..... | 79 |

LISTA DE TABLAS

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Tabla 1. Matriz de errores de orden $k \times k$ | 31 |
| Tabla 2. Estadísticas descriptivas de la RSP (Mpa) en la capa 0.0 - 0.10 m de profundidad .48 | |
| Tabla 3. Informaciones sobre el semivariograma omnidimensional muestral de Matheron para la RSP (Mpa) en la capa 0.0-0.10 m de profundidad | 50 |
| Tabla 4. Análisis de la estructura de dependencia espacial para la variable RSP, utilizando el método de estimación ML | 51 |
| Tabla 5. Criterios de validación cruzada, Akaike y Schwarz en la selección del mejor ajuste | 51 |
| Tabla 6. Estadísticas descriptivas de la RSP (Mpa) en la capa 0.0-0.10 m de profundidad sin el punto 01 considerado influyente | 55 |
| Tabla 7. Análisis de la estructura de dependencia espacial para la RSP sin el punto 01 considerado influyente, utilizando o método de estimación de parámetros ML..... | 56 |
| Tabla 8. Criterios de validación cruzada, Akaike- A/C y Schwarz- S/C para seleccionar el mejor modelo ajustado..... | 56 |
| Tabla 9. Distribución del número de píxeles y porcentaje de área por clases para el mapa con todos los puntos | 57 |
| Tabla 10. Distribución del número de píxeles y porcentajes de área por clases para el mapa sin observación influyente | 57 |
| Tabla 11. Matriz de errores para la comparación de mapas temáticos de la RSP en la capa 0 - 0.10 m de profundidad, por número de <i>píxeles</i> | 58 |
| Tabla 12. Estadísticas descriptivas de la Productividad de trigo (PROD) ($t\ ha^{-1}$) total y por variedades CD 101 y CD 103 | 59 |
| Tabla 13. Modelos ajustados por ML, estimativas de los parámetros y desviaciones estándares (entre paréntesis) para el estudio de la variabilidad espacial de la Productividad de trigo..... | 62 |
| Tabla 14. Criterios de validación cruzada, A/C y S/C para la selección del mejor modelo ajustado para la Productividad de trigo..... | 62 |
| Tabla 15. Modelos ajustados por ML y estimaciones de los parámetros del modelo espacial lineal con sus desviaciones estándares (entre paréntesis), para la Productividad de trigo | 66 |
| Tabla 16. Índices de acuracia: exactitud global y kappa para la comparación de mapas temáticos de la Productividad de trigo con y sin los puntos influyentes | 68 |

LISTA DE FIGURAS

| | |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figura 1. Relación entre la función de covarianza espacial $C(h)$ y la función semivarianza $\gamma(h)$. | 10 |
| Figura 2. Gráfico de un semivariograma. | 11 |
| Figura 3. Función semivarianza $\gamma(h)$ en las direcciones 0° , 45° , 90° y 135° . | 12 |
| Figura 4. Gráficos de los modelos esférico (a), exponencial (b), gaussiano (c) para la función semivarianza $\gamma(h)$ y gráficos de los modelos exponencial, gaussiano y de la familia Matérn con parámetro de orden 0.7 para las función de covarianza $C(h)$ (d). | 16 |
| Figura 5. Esquema muestral del área I en estudio con una cuadrícula regular de 75 x 75 m. | 45 |
| Figura 6. Esquema muestral del área II en estudio, con una cuadrícula regular de 50 x 50 m. | 46 |
| Figura 7. Gráficos <i>boxplot</i> de la RSP en la capa 0.0 - 0.10 m de profundidad. | 49 |
| Figura 8. Gráfico <i>postplot</i> de la RSP en la capa 0.0 - 0.10 m de profundidad. | 49 |
| Figura 9. Gráfico de los semivariogramas direccionales de la RSP (a); Gráfico del semivariograma omnidimensional muestral de la RSP (b). | 50 |
| Figura 10. Mapa temático de la RSP [Mpa] en la capa 0.0 - 0.10 m de profundidad. | 52 |
| Figura 11. Gráfico C_i versus a orden, de la RSP utilizando la perturbación aditiva | 53 |
| Figura 12. Gráfico $ L_{max} $ versus orden, de la RSP utilizando la perturbación aditiva. | 53 |
| Figura 13. Gráfico C_i versus a orden, de la RSP, utilizando la perturbación de Zhu. | 53 |
| Figura 14. Gráfico $ L_{max} $ versus orden, de la RSP, utilizando la perturbación de Zhu. | 53 |
| Figura 15. Gráfico C_i versus a orden, de la RSP utilizando la perturbación de Zhu generalizada. | 54 |
| Figura 16. Gráfico $ L_{max} $ versus orden, de la RSP utilizando la perturbación de Zhu generalizada. | 54 |
| Figura 17. Gráfico <i>postplot</i> de la RSP en la capa 0.0 - 0.10 m de profundidad sin la observación 01 considerado influyente. | 55 |
| Figura 18. Mapa temático de la RSP en la capa 0.0 - 0.10 m de profundidad sin la observación 01 considerada influyente. | 57 |
| Figura 19. Gráfico <i>boxplot</i> (a); gráfico <i>postplot</i> (b) para los datos de la Productividad de trigo. | 60 |
| Figura 20. Gráficos de las funciones semivarianza muestral en las direcciones de 0° , 45° , 90° y 135° (a); semivariograma omnidimensional muestral (b). | 61 |
| Figura 21. Mapa temático de la Productividad de trigo por áreas, considerando la variedad CD 101 en la subárea 1 y la variedad CD 103 en las subáreas 2 y 3. | 63 |
| Figura 22. Gráfico $D_{i\theta}$ versus orden de los datos (a); Gráfico $D_{i\beta}$ versus orden de los datos (b); Gráfico $D_{i\varphi}$ versus orden de los datos (c) para el estudio de influencia global. | 64 |

| | |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figura 23. Gráficos de influencia local B_i versus orden (a) y l_{max} versus orden (b) para la perturbación aditiva; Gráficos de influencia local B_i versus orden (c) y $ L_{max} $ versus orden (d) para la perturbación de Zhu; Gráficos de influencia local B_i versus orden (e) y $ L_{max} $ versus orden (f) para la perturbación de Zhu generalizada..... | 65 |
| Figura 24. Mapa de la Productividad de trigo por áreas con todos los datos (a); Mapa Productividad de trigo sin la observación 08 (b); Mapa de Productividad de trigo sin la observación 57 (c); Mapa de Productividad de trigo sin la observación 08 y 57 (d); Mapa da Productividad de trigo sin la observación 09 (e) y Mapa de Productividad de trigo sin las observaciones 09 y 57 (f) consideradas influyentes..... | 67 |

INTRODUCCIÓN

La geoestadística es un importante método de análisis de datos para el estudio de la variabilidad espacial (dependencia espacial) de variables aleatorias en un espacio georreferenciado. Ella surge para se referir a la aplicación de la teoría de las variables regionalizadas en la solución de problemas de minería y geología. La geoestadística fue idealizada por Matheron entre los años de 1957 a 1963, es un criterio científico y moderno en el estudio de la estructura de dependencia espacial de diversos fenómenos naturales.

MATHERON (1963) denominada variable regionalizada a una variable aleatoria continua que es espacialmente georreferenciada y la define como una función espacial que varía de un lugar para otro, con una continuidad.

La geoestadística fue influenciada por los trabajos estadísticos en minería de la Escuela de Sudáfrica, a cargo de Krige y Sichei en el año 1951 y ha sido ampliamente utilizada en los últimos años en diferentes fenómenos naturales, en el estudio de zonas forestales, en el estudio de variables agrometeorológicas, geología, minería, agronomía, en las ciencias de la tierra, en el estudio de la productividad de cultivos; en el estudio de característica físicas y químicas del suelo, tales como: concentración de nutriente del suelo, niveles de materia orgánica, conductividad hidráulica, etc.

La geoestadística tiene como principio que los elementos de la muestra de una variable regionalizada están dispuestos de forma correlacionada hasta cierta distancia y tienen influencia sobre los puntos próximos recolectados que serán estimados. La geoestadística es utilizada cuando se comprueba la existencia de dependencia espacial entre elementos de la muestra, teniendo en consideración las distancias entre los elementos de la muestra y la autocorrelación entre ellos.

Las variables regionalizadas tienen sus características relacionadas con la estructura del fenómeno natural, la cual ella representa. Estas características son (CRESSIE, 1993): ubicación, continuidad espacial y la ausencia de direcciones privilegiadas (anisotropía).

En el estudio de la estructura de dependencia espacial de una variable regionalizada son utilizadas la función semivarianza, la función de covarianza o la función de autocorrelación espacial, que permiten representar la dependencia espacial entre unidades de muestreo vecinas, utilizando modelos espaciales. Del modelo espacial son obtenidas las estimaciones de los parámetros que caracterizan las medidas de correlación existentes entre los valores tomados en dos puntos en el espacio y del radio de dependencia espacial entre las unidades de muestreo.

La geoestadística utiliza la técnica de interpolación denominada kriging para la construcción de mapas temáticos, la cual usa la estructura de dependencia espacial obtenida mediante el ajuste de un modelo espacial para la estimación de valores en lugares no muestreados, teniendo estimaciones sin sesgo y con varianza mínima. Estas dos características garantizan que la estimaciones obtenidas por la interpolación por kriging nos proporcionan el mejor estimador lineal insesgado de varianza mínima - BLUE (*Best Linear Unbiased Estimator*) (WEBSTER & OLIVER, 2007).

Los mapas temáticos contruidos por los métodos geoestadísticos son excelentes recursos para el análisis del desempeño de las variables sobre la región en estudio. Sin embargo, cuando la serie de datos tiene algunos valores atípicos pueden causar interferencia desproporcionada en el estudio geoestadístico, cambiando la elección del modelo ajustado, alterando las estimaciones de los parámetros que definen la estructura de dependencia espacial y por lo tanto, alterando la construcción de los mapas temáticos por la técnica kriging, cambiando la interpretación de los resultados obtenidos; sufriendo distorsiones de la realidad y en consecuencia puede causar una interpretación errónea.

Observaciones que afectan las decisiones de elección, las estimaciones de los parámetros y de las estimaciones de los valores en lugares no muestreados se denominan puntos

influyentes. Es importante destacar que existen puntos influyentes que por el análisis estadístico tradicional no se consideran atípicos, sin embargo, pueden ser detectados mediante los diagnósticos de influencia.

En el estudio de métodos de diagnóstico de puntos influyentes, la literatura presenta diagnósticos de influencia global y de influencia local. El diagnóstico de influencia global fue propuesto por COOK (1977), y se basa en la eliminación individual de puntos del conjunto total de datos. Sin embargo, un problema que puede ocurrir con la eliminación de las observaciones individuales es la no detección de puntos conjuntamente influyentes. Además, cuanto mayor es la cantidad de datos recolectados, se requiere de mayor costo computacional. Para reducir estos problemas y optimizar el análisis, COOK (1986) presenta una sugerencia considerada innovadora en el estudio de métodos de diagnóstico llamado Influencia local, que es evaluar la influencia conjunta de las observaciones bajo pequeñas perturbaciones en el modelo, en lugar de la evaluación realizada retirando puntos.

La metodología de influencia local (COOK, 1986; OSORIO, 2006; OSORIO et al., 2007; PAULA, 2013) se ha convertido en una técnica de diagnóstico popular para la identificación conjunta de observaciones influyentes en los modelos de regresión, con la ventaja de aplicarse a cualquier modelo paramétrico. En los últimos años, los diagnósticos de influencia local han sido propuestos por varios autores. GALEA et al. (2003, 2005) llevaron a cabo estudios de influencia local en los modelos de regresión lineal y no lineal para muestras independientes, con la presentación de diversas formas de perturbación. En el estudio de los diagnósticos para las variables regionalizadas, CHRISTENSEN et al. (1993); MILITINO et al. (2004) estudiaron los métodos de influencia global, basadas en la eliminación de uno o más puntos, para verificar el efecto de los puntos de influencia en las estimaciones de los parámetros. BORSSOI et al. (2009) estudiaron los métodos de influencia local, para verificar el efecto de los puntos influyentes en las estimaciones de los parámetros de máxima verosimilitud y la construcción de mapas temáticos mediante la técnica kriging; ASSUMPÇÃO et al. (2011) presentaron las técnicas para influencia local en los datos con distribución t-

Student con grados de libertad fijos, teniendo en cuenta la perturbación aditiva sobre la variable respuesta y URIBE-OPAZO et al. (2012) estudiaron las técnicas de análisis de influencia local para comprobar la sensibilidad de las estimaciones de máxima verosimilitud (ML) y de máxima verosimilitud restringida (MLR) en presencia de puntos influyentes en modelos espaciales lineales gaussianos, cuando pequeñas perturbaciones aditivas son colocadas en la variable respuesta.

Este trabajo tiene como objetivo presentar los métodos geoestadísticos en el estudio de la variabilidad espacial de variables regionalizadas utilizando los modelos espaciales lineales gaussianos; estudiar las técnicas de diagnóstico de influencia global y local de las variables regionalizadas utilizando tres esquemas de perturbación para la variable respuesta que son presentadas en la literatura. Finalmente, la metodología presentada es aplicada en dos conjuntos de datos reales, evaluando la metodología desarrollada.

La tesis es dividida de la siguiente manera:

En el Capítulo 1, se presenta una revisión de los conceptos de geoestadística, estimación de parámetros, kriging y criterios de selección de modelos.

En el Capítulo 2, se presenta una comparación de mapas temáticos, a través de los índices de exactitud.

En el Capítulo 3, se estudia el modelo espacial lineal, la función score, la matriz de información observada e esperada de Fisher.

En el Capítulo 4, se presenta el estudio de las técnicas de diagnóstico para influencia global e de influencia local en los modelos espaciales lineales.

En el Capítulo 5, se presenta la teoría descrita para dos conjuntos de datos reales.

En el Capítulo 6, se presentan los resultados y discusiones de los análisis de los dos conjuntos de datos.

En el Capítulo 7, son presentadas las conclusiones del trabajo.

CAPÍTULO 1

GEOESTADÍSTICA

La geoestadística tiene como base la teoría de las variables regionalizadas, que es el estudio de las funciones aleatorias que varían de un lugar a otro en el espacio con una cierta continuidad, esto es, de variables aleatorias cuyos valores están de alguna manera relacionados con la posición que ocupan en el espacio. Esta teoría afirma que algunas medidas pueden ser vistas como una realización de una función aleatoria o proceso estocástico (MATHERON, 1963).

La teoría de las variables regionalizadas desarrollada por Matheron tiene en cuenta que las variables aleatorias están autocorrelacionadas hasta una cierta distancia, ya que los valores numéricos pueden variar considerablemente de un punto a otro en el espacio.

La geoestadística permite definir la estructura de dependencia espacial teniendo en cuenta la ubicación, la continuidad espacial y la isotropía de los datos (no hay una dirección privilegiada en el estudio de la autocorrelación espacial), también permite obtener el radio de dependencia espacial entre las unidades de muestreo, así como la construcción de mapas temáticos interpolados considerando los datos muestreados y la estructura de dependencia espacial obtenida. Los elementos de la muestra son recolectados en lugares específicos, ubicados por sus coordenadas geográficas, como por ejemplo, el sistema de coordenadas *Universal Transverse Mercator* (UTM).

La continuidad espacial presentada entre las unidades de muestreo vecinas, refleja el grado de dependencia espacial. Cuando la dependencia espacial no es identificada, se entiende que hay presencia de un efecto llamado de “pepita puro”, esto es, no hay autocorrelación espacial entre elementos de la muestra o la distancia entre los elementos de la muestra puede haber sido demasiado grande como para no identificar la dependencia espacial de la variable en estudio.

1.1 Teoría de las variables regionalizadas

MARDIA & MARSHALL (1984) consideran un proceso estocástico estacionario $\{Z(\mathbf{s}_i), \mathbf{s}_i \in \mathcal{S}\}$, donde $\mathcal{S} \subset \mathbb{R}^2$, siendo \mathbb{R}^2 el espacio euclidiano bidimensional. El proceso estocástico tiene un conjunto de variables aleatorias $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))^T$ que representan las variables regionalizadas, siendo $\mathbf{s}_i, i = 1, \dots, n$, los lugares espaciales conocidas.

Sea \mathbf{Z} el vector aleatorio que tiene una distribución normal n -variada con vector de medias $\mu\mathbf{1}$ y matriz de covarianza Σ , esto es, $\mathbf{Z} \sim N_n(\mu\mathbf{1}, \Sigma)$, donde, μ es una constante; $\mathbf{1}$ es un vector de unos $n \times 1$ y Σ es una matriz $n \times n$, no singular, simétrica y definida positiva, dada por $\Sigma = [(\sigma_{iu})]$, donde, $\sigma_{iu} = \text{Cov}[Z(\mathbf{s}_i), Z(\mathbf{s}_u)]$, para $i, u = 1, \dots, n$.

La variación espacial de una variable regionalizada $Z(\mathbf{s}_i)$ puede ser expresada por una suma de dos componentes: una componente estructural $\mu(\mathbf{s}_i)$, asociado a un valor medio constante y un componente aleatorio $\varepsilon(\mathbf{s}_i)$, espacialmente correlacionado, descritos por el modelo de la ecuación (1):

$$Z(\mathbf{s}_i) = \mu(\mathbf{s}_i) + \varepsilon(\mathbf{s}_i), \quad (1)$$

donde, $\mu(\mathbf{s}_i)$ y $\varepsilon(\mathbf{s}_i)$ pueden depender de la localización espacial en la cual $Z(\mathbf{s}_i)$ fue obtenida. Si asumimos que el error aleatorio $\varepsilon(\mathbf{s}_i)$ tiene media cero, esto es, $E[\varepsilon(\mathbf{s}_i)] = 0$, y que la variación entre puntos en el espacio es determinada por alguna función de covarianza $\text{Cov}[\varepsilon(\mathbf{s}_i), \varepsilon(\mathbf{s}_u)] = \text{Cov}[Z(\mathbf{s}_i), Z(\mathbf{s}_u)] = C(\mathbf{s}_i, \mathbf{s}_u) = \sigma_{iu}$, para $i, u = 1, \dots, n$.

El modelo espacial depende de la estructura de la matriz de covarianza Σ del proceso estocástico gaussiano \mathbf{Z} , teniendo en cuenta una forma paramétrica de manera particular dado por MARDIA & MARSHALL (1984), en la ecuación (2):

$$\Sigma = \varphi_1 \mathbf{I}_n + \varphi_2 \mathbf{R}(\varphi_3), \quad (2)$$

donde, \mathbf{I}_n es la matriz identidad, $n \times n$, el parámetro $\varphi_1 \geq 0$ es denominado de “efecto pepita” (*nugget*); el parámetro $\varphi_2 \geq 0$ es denominada la contribución (*sill*); el parámetro φ_3 es un parámetro que determina el radio de dependencia espacial. $\mathbf{R}(\varphi_3) = [(r_{iu})]$, es una matriz simétrica $n \times n$, que es función del parámetro $\varphi_3 > 0$. Esta matriz tiene como elementos en la

diagonal principal $r_{ii} = 1$ y fuera de la diagonal principal los elementos $r_{iu} = \varphi_2^{-1} C(s_i, s_u)$, para $\varphi_2 \neq 0$ y $r_{iu} = 0$, para $\varphi_2 = 0$, $i \neq u = 1, \dots, n$, siendo que r_{iu} depende de la distancia euclidiana $h_{iu} = \|s_i - s_u\|$, entre los puntos s_i y s_u .

La forma paramétrica de la matriz de covarianza Σ , representada en la ecuación 2, ocurre para varios procesos estacionarios isotrópicos, en la cual la covarianza $C(s_i, s_u) = C(h_{iu})$ es definida como $C(h_{iu}) = \varphi_2 r_{iu}$. En la función de covarianza $C(h_{iu})$, la varianza del proceso estocástico \mathbf{Z} es dada por $\text{Var}(Z(\mathbf{s}_i)) = C(0) = \varphi_1 + \varphi_2$.

Con una única muestra de tamaño n teniendo un único elemento muestral en cada punto s_i , todo lo que se conoce de un proceso estocástico \mathbf{Z} es una única realización. Entonces, para estimar valores para los locales no muestreados, se tiene que introducir la restricción de que la variable regionalizada Z sea necesariamente estacionaria de segundo orden, esto es, que los momentos estadísticos hasta segunda orden de la variable aleatoria $Z(\mathbf{s} + \mathbf{h})$ sean los mismos para cualquier valor \mathbf{h} .

Supongamos que el proceso estocástico \mathbf{Z} es estacionario de segundo orden, esto es, el valor esperado $E[Z(\mathbf{s})] = E[Z(\mathbf{s} + \mathbf{h})] = \mu$, con la misma variancia y la covarianza entre $Z(\mathbf{s})$ y $Z(\mathbf{s} + \mathbf{h})$ es definida por la ecuación (3) (COX & HINKEY, 1974):

$$\text{Cov}(Z(\mathbf{s}), Z(\mathbf{s} + \mathbf{h})) = E[Z(\mathbf{s})Z(\mathbf{s} + \mathbf{h})] - E[Z(\mathbf{s})]E[Z(\mathbf{s} + \mathbf{h})] = C(\mathbf{h}), \quad (3)$$

la función $2\gamma(\mathbf{s}, \mathbf{s} + \mathbf{h})$ es definida por

$$2\gamma(\mathbf{s}, \mathbf{s} + \mathbf{h}) = 2\gamma(\mathbf{h}) = E[(Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{h}))^2].$$

De esta manera, se define la función semivarianza $\gamma(\mathbf{h})$ como:

$$\gamma(\mathbf{h}) = \frac{1}{2} E[(Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{h}))^2], \quad (4)$$

tal que $\|\mathbf{h}\| = h$ é a distância euclidiana entre pontos (*lag*).

La relación entre la función semivarianza $\gamma(\mathbf{h})$ y la función de covarianza $C(\mathbf{h})$ es dada como:

$$\gamma(\mathbf{h}) = C(0) - C(\mathbf{h}). \quad (5)$$

La función semivarianza $\gamma(\mathbf{h})$ tiene las siguientes propiedades:

- i) $\gamma(0) = 0$;
- ii) $\gamma(h) = \gamma(-h) \geq 0$;
- iii) $\lim_{n \rightarrow \infty} \frac{\gamma(h)}{|h|^2} = 0$, esto es, $\gamma(h)$ crece mas lentamente que h^2 .

A partir de la función semivarianza presentada en la ecuación (5), se tiene que para $C(0) \neq 0$ la relación $\frac{C(h)}{C(0)} = 1 - \frac{\gamma(h)}{C(0)}$. De esta manera, la función de autocorrelación espacial $\rho(h)$ es de la forma dada en (6):

$$\rho(h) = 1 - \frac{\gamma(h)}{C(0)}. \quad (6)$$

Por tanto, si la hipótesis de estacionariedad de segundo orden es satisfecha, la covarianza $C(h)$ y la función semivarianza $\gamma(h)$ caracterizan la dependencia espacial (ISAACS & SRIVASTAVA, 1989).

La existencia de estacionariedad nos ofrece la oportunidad de repetir un experimento aun cuando los elementos de la muestra puedan haber sido recolectados en puntos diferentes, porque todos ellos son consideradas que pertenecen a una misma población con los mismos parámetros.

Un estimador de la función semivarianza definida en la ecuación (4), es conocido como el estimador de los momentos o estimador de Matheron y es definido como:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [(Z(s_i) - Z(s_i + h))^2], \quad (7)$$

donde, $N(h)$ es el número de pares $Z(s_i) - Z(s_i + h)$, separados por vectores h , tal que $\|h\| = h$ é a distância euclidiana entre pontos ("lag").

1.2 Semivariograma

El semivariograma es el gráfico de la función semivarianza $\gamma(h)$ versus la distancia h . Este gráfico es utilizado para analizar la estructura de dependencia espacial entre los elementos de la muestra dentro de un espacio paramétrico, siendo que esta estructura de dependencia espacial es obtenida a través de un modelo teórico ajustable a la función semivarianza (o

función covarianza) y las estimaciones de los parámetros del modelo ajustado son obtenidas a partir de los datos de la muestra. Este modelo teórico ajustado será utilizado posteriormente en la interpolación espacial de valores en lugares no muestreados, por medio de la técnica de kriging.

Sea $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \varphi_3)^T$ el vector de parámetros que define la estructura de dependencia espacial, presentada en la ecuación (2), que tiene como características:

- Idealmente se tiene que $\gamma(0) = 0$, esto es, en la teoría cuando la distancia h es cero, el valor de la función semivarianza es igual a cero. En la práctica, a medida que h tiende a cero, $\gamma(h)$ se aproxima al valor φ_1 llamado efecto pepita, que revela la discontinuidad del semivariograma para distancias menores que la menor distancia entre los elementos de la muestra. CRESSIE (1993) interpreta el efecto pepita φ_1 como un error de medición o de variabilidad en pequeña escala;

- La contribución (φ_2) es conocida como *sill*, que es la diferencia entre la meseta (C) y el efecto pepita (φ_1);

- La meseta ($C = \varphi_1 + \varphi_2$) es el valor en el cual la función semivarianza $\gamma(h)$ se estabiliza. Desde este punto hacia adelante, se considera que no existe dependencia espacial entre los elementos muestrales, porque la varianza de la diferencia entre pares ($VAR [Z(s_i) - Z(s_i + h)]$) es invariante con la distancia;

- El alcance o rango (range) $a = g(\varphi_3)$, es una función del parámetro φ_3 , donde g representa una función positiva. Es la distancia dentro de la cual los elementos de la muestra son auto correlacionados espacialmente. De esta manera, el alcance determina la distancia a partir de la cual la variable regionalizada deja de ser e autocorrelacion.

En la Figura 1 se presenta la función de covarianza espacial $C(h)$, la función semivarianza $\gamma(h)$ y la relación entre estas dos funciones presentada en la ecuación (5).

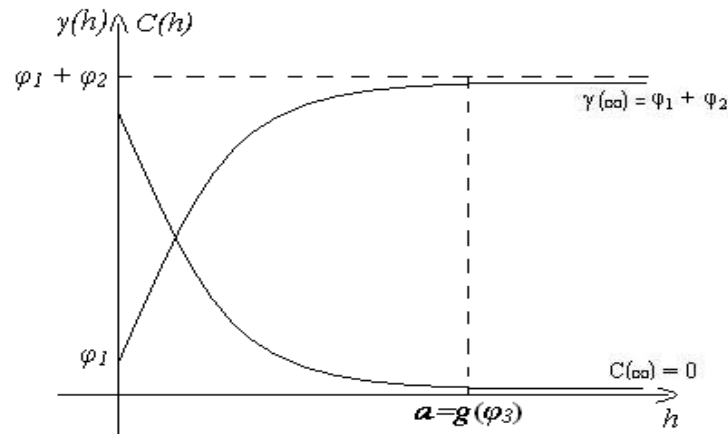


Figura 1. Relación entre la función de covarianza espacial $C(h)$ y la función semivarianza $\gamma(h)$.

Para la construcción de un semivariograma a demás de determinar el estimador para la función semivarianza es importante definir la distancia máxima hasta donde será construido. CLARK (1979) comenta que un semivariograma debe ser construido hasta 50% de la distancia máxima entre los puntos de la muestra (*cutoff*), el autor comenta que después de esta distancia la semivarianza pierde el significado, pero, nada impide que sea usadas distancias mayores, siempre y cuando tenga sentido la elección.

En la construcción del semivariograma es necesario determinar el número de puntos h (*lags*) para los cuales las semivarianzas experimentales serán estimadas. Si el plano de muestreo es regular, la propia distancia entre puntos determina los “*lags*”, en este caso, cada semivarianza experimental puede ser calculada con base en todos los pares de puntos de la muestra separados por un mismo vector h . En el caso de tener un plano de muestreo irregular es posible que se tenga un número pequeño de pares de puntos separados por el mismo vector h . McBRATNEY & WEBSTER (1986) recomiendan que sea admitida una tolerancia Δh , en la distancia y una tolerancia angular ϑ en la dirección, de tal forma que satisfaga el requisito de tener el número mínimo de puntos para la estimación de la función semivarianza a cada “*lag*”. Tanto para la distancia h , como para la tolerancia angular ϑ , son necesarios intervalos de variación. La distancia h depende del tamaño del área experimental en estudio y

del punto de corte ("*cutoff*") determinado. Para la tolerancia angular, ISAAKS & SRIVASTAVA (1989) proponen una tolerancia de $\vartheta = 40^\circ$.

El número de pares con que será calculada la función semivarianza ($N(h)$) desempeña un papel importante en la construcción del semivariograma, siendo necesario según JOURNEL & HUIJBREGTS (1978) y WOLLENHAUPT *et al.* (1997), de un mínimo de 30 pares.

La Figura 2 ilustra un semivariograma con características próximas del ideal. El patrón de comportamiento de la función semivarianza, que se espera es que las diferencias $[Z(s) - Z(s + h)]$ crezcan a medida que la distancia que los separa (h) aumente hasta estabilizarse. Se espera que observaciones más próximas geográficamente tengan un comportamiento más semejante entre sí que aquellas separadas por mayores distancias. De esta manera, se espera que la función semivarianza $\gamma(h)$ aumente con la distancia h hasta alcanzar un valor máximo en la cual se estabilice.

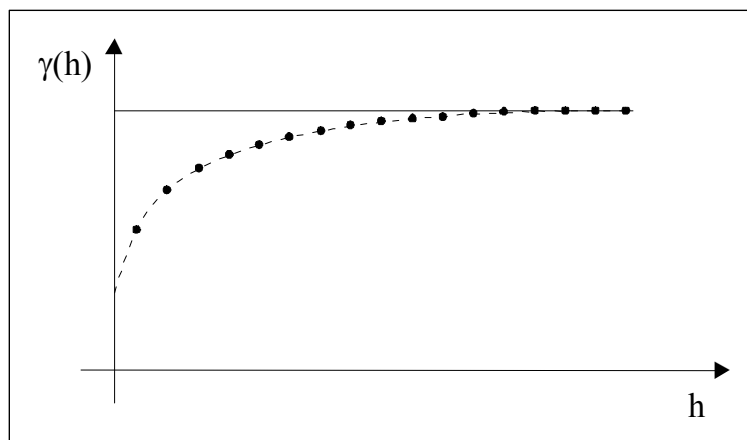


Figura 2. Gráfico de un semivariograma.

Si el semivariograma en vez de crecer se mantiene constante a medida que la distancia h aumenta, entonces se tiene un efecto llamado de pepita puro y define la falta de dependencia espacial. Esto significa que el alcance a , para los datos en estudio, es menor que el menor espacio entre elementos de la muestra. En esta situación se tiene una distribución de datos no correlacionados.

1.3 Anisotropía

Cuando los semivariogramas construidos en diferentes direcciones se presentan con los mismos comportamientos, sugiriendo el mismo padrón de continuidad espacial, el fenómeno en estudio es considerado isotrópico (Figura 3). En esta situación, el estudio de dependencia espacial puede ser realizado a través de un único semivariograma, denominado semivariograma omnidireccional, o semivariograma isotrópico, en que las semivarianzas son estimadas con información de todas las direcciones, tomando en consideración solo el módulo del vector h y no la dirección. La isotropía es de gran importancia para la estimación de valores de una variable para lugares no muestreados, pues, ella indica que el área de influencia de cada valor de la muestra es circular, determinando la forma de la ventana de estimación a ser utilizada en la interpolación por kriging (GUEDES et al., 2008).

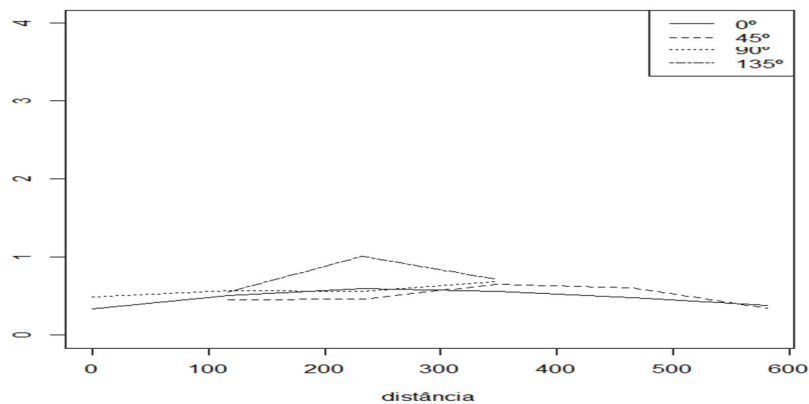


Figura 3. Función semivarianza $\gamma(h)$ en las direcciones 0°, 45° 90° y 135°.

Cuando los semivariogramas construidos en diferentes direcciones no tienen los mismos comportamientos, el fenómeno en estudio es considerado anisotrópico. Según CLARK (1979); WACKERNAGEL (2003) y GUEDES et al. (2008) la anisotropía es encontrada a través de la existencia de semivariogramas, que construidos en diferentes direcciones, revelan una privilegiada dirección con distinto comportamiento de dependencia espacial. CÂMARA & MEDEIROS (1998) describen que la anisotropía puede ser fácilmente detectada por medio de

la construcción y observación de los semivariogramas en distintas direcciones. Las direcciones definidas como convencionales en estudios geoestadísticos son: 0° , 45° , 90° y 135° , pudiéndose trabajar con otras direcciones (ISAAKS & SRIVASTAVA, 1989; GUEDES et al., 2008).

Los principales tipos de anisotropía encontradas en los estudios son: anisotropía geométrica, anisotropía zonal y anisotropía combinada. La anisotropía geométrica se caracteriza porque los semivariogramas presentan el mismo modelo, con la misma meseta en todas las direcciones, pero con diferentes alcances, verificándose los alcances máximos y mínimos en direcciones perpendiculares (ejemplo: direcciones 0° y 90°). La anisotropía zonal se caracteriza porque los semivariogramas, contruidos para diferentes direcciones, presentan el mismo modelo, con diferentes mesetas. Finalmente, la anisotropía combinada se caracteriza por ser la combinación de la anisotropía geométrica y zonal. (ISAAKS & SRIVASTAVA, 1989). Es importante destacar que en el caso de la existencia de anisotropía, esta debe ser incorporada en el modelo en estudio a través de un factor de anisotropía. (GUEDES et al., 2008).

1.4 Modelos espaciales

En esta sección se presentan algunos de los modelos espaciales teóricos para el ajuste de la función semivarianza, la función de covarianzas o la función de autocorrelación, que son divididos en modelos transitivos los que tienen meseta y no transitivos los que no tienen meseta (ISAAKS & SRIVASTAVA, 1989; CRESSIE, 1993). Los modelos transitivos más utilizados son: esférico, exponencial, gausiano y familia Matérn (BARNES, 1991; WESTER & OLIVER, 2007), estos modelos consideran que la variable regionalizada tiene un mismo comportamiento de continuidad espacial en todas las direcciones (isotropía), de modo que el semivariograma debe tener la misma característica estructural en todas las direcciones del espacio.

A Continuación se presenta los modelos espaciales: esférico, exponencial, gaussiano y familia Matérn, válidos en \mathbb{R} , \mathbb{R}^2 y \mathbb{R}^3 (ISAAKS & SRIVASTAVA, 1989; CRESSIE, 1993).

i) Modelo esférico: presenta crecimiento rápido en el origen y alcanza la meseta a 2/3 del alcance. La función semivarianza teórica tiene como expresión la ecuación:

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ \varphi_1 + \varphi_2 \left[\frac{3}{2} \left(\frac{h}{\varphi_3} \right) - \frac{1}{2} \left(\frac{h}{\varphi_3} \right)^3 \right], & 0 < h \leq \varphi_3 \\ \varphi_1 + \varphi_2, & h > \varphi_3. \end{cases} \quad (8)$$

La función de covarianza teórica se expresa como la ecuación:

$$C(h) = \begin{cases} \varphi_1 + \varphi_2, & h = 0 \\ \varphi_2 \left[1 - \frac{3}{2} \left(\frac{h}{\varphi_3} \right) + \frac{1}{2} \left(\frac{h}{\varphi_3} \right)^3 \right], & 0 < h \leq \varphi_3 \\ 0, & h > \varphi_3. \end{cases} \quad (9)$$

La función de autocorrelación espacial teórica es definida por la ecuación:

$$\rho(h) = \begin{cases} 1, & h = 0 \\ 1 - \frac{3}{2} \left(\frac{h}{\varphi_3} \right) + \frac{1}{2} \left(\frac{h}{\varphi_3} \right)^3, & 0 < h \leq \varphi_3. \end{cases} \quad (10)$$

ii) Modelo exponencial: presenta comportamiento aproximadamente lineal en el origen y alcanza la meseta asintóticamente, con alcance práctico definido como la distancia en la cual el valor del modelo es 95% de φ_2 , siendo que el alcance es dado por $a = 3\varphi_3$. La función semivarianza teórica tiene como expresión la ecuación:

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ \varphi_1 + \varphi_2 \left[1 - \exp \left(-\frac{h}{\varphi_3} \right) \right], & h > 0. \end{cases} \quad (11)$$

La función de covarianza teórica es expresada por la ecuación

$$C(h) = \begin{cases} \varphi_1 + \varphi_2, & h = 0 \\ \varphi_2 \left[\exp \left(-\frac{h}{\varphi_3} \right) \right], & h > 0. \end{cases} \quad (12)$$

La función de autocorrelación espacial teórica es definida por la ecuación:

$$\rho(h) = \begin{cases} 1, & h = 0 \\ \left[\exp \left(-\frac{h}{\varphi_3} \right) \right], & h > 0. \end{cases} \quad (13)$$

iii) Modelo gausiano: presenta comportamiento parabólico en el origen y es utilizado para modelar un fenómeno extremadamente continuo. También alcanza la meseta asintóticamente y el alcance es dado por $a = \sqrt{3} \varphi_3$. La función semivarianza teórica del modelo gausiano tiene como expresión la ecuación:

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ \varphi_1 + \varphi_2 \left\{ 1 - \exp \left[- \left(\frac{h}{\varphi_3} \right)^2 \right] \right\}, & h > 0. \end{cases} \quad (14)$$

La función de covarianza teórica es expresada por la ecuación:

$$C(h) = \begin{cases} \varphi_1 + \varphi_2, & h = 0 \\ \varphi_2 \left\{ \exp \left[- \left(\frac{h}{\varphi_3} \right)^2 \right] \right\}, & h > 0. \end{cases} \quad (15)$$

La función de autocorrelación espacial teórica es definida por la ecuación:

$$\rho(h) = \begin{cases} 1, & h = 0 \\ \exp \left[- \left(\frac{h}{\varphi_3} \right)^2 \right], & h > 0. \end{cases} \quad (16)$$

iv) Familia Matérn: presenta por MATÉRN, (1986), es una función positiva definida en términos del modelo teórico de semivarianza, dada como:

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ \varphi_1 + \varphi_2 \left[1 - \left(2^{k-1} \Gamma(k) \right)^{-1} \left(\frac{h}{\varphi_3} \right)^k K_k \left(\frac{h}{\varphi_3} \right) \right], & h > 0, \end{cases} \quad (17)$$

donde, $k > 0$, $\Gamma(k) = \int_0^\infty x^{k-1} e^{-x} dx$ es la función gama y $K_k(u) = \frac{1}{2} \int_0^\infty x^{k-1} e^{-\frac{1}{2}u(x+x^{-1})} dx$,

es la función de Bessel modificada de tercer tipo de orden k (GRADSHTEYN & RYZHIK, 2000). En la familia Matérn, el parámetro de orden k es un parámetro de forma que determina la suavización del proceso subyacente \mathbf{Z} .

La función de covarianza teórica de la Familia Matérn para la orden $k > 0$ fijo, es dada por la ecuación:

$$C(h) = \begin{cases} \varphi_1 + \varphi_2, & h = 0 \\ \varphi_2 \left[\left(2^{k-1} \Gamma(k) \right)^{-1} \left(\frac{h}{\varphi_3} \right)^k K_k \left(\frac{h}{\varphi_3} \right) \right], & h > 0. \end{cases} \quad (18)$$

La función de autocorrelación espacial de la Familia Matérn para el orden $k > 0$, fijo, es definida por la ecuación:

$$\rho(h) = \{2^{k-1}\Gamma(k)\}^{-1} \left(\frac{h}{\varphi_3}\right)^k K_k\left(\frac{h}{\varphi_3}\right), \quad h > 0. \quad (19)$$

La familia Matérn corresponde al modelo exponencial cuando el parámetro de ordenes $k = 0.5$; modelo de Whittle cuando $k = 1$ y modelo gaussiano cuando $k \rightarrow \infty$ (WEBSTER & OLIVER, 2007).

En la Figura 4 son presentados ejemplos de los gráficos de los modelos esférico, exponencial, gaussiano para la función semivarianza $\gamma(h)$ y de la familia Matérn con parámetro de ordenes $k = 0.5, 0.7$ y $k \rightarrow \infty$, para la función de covarianza $C(h)$.

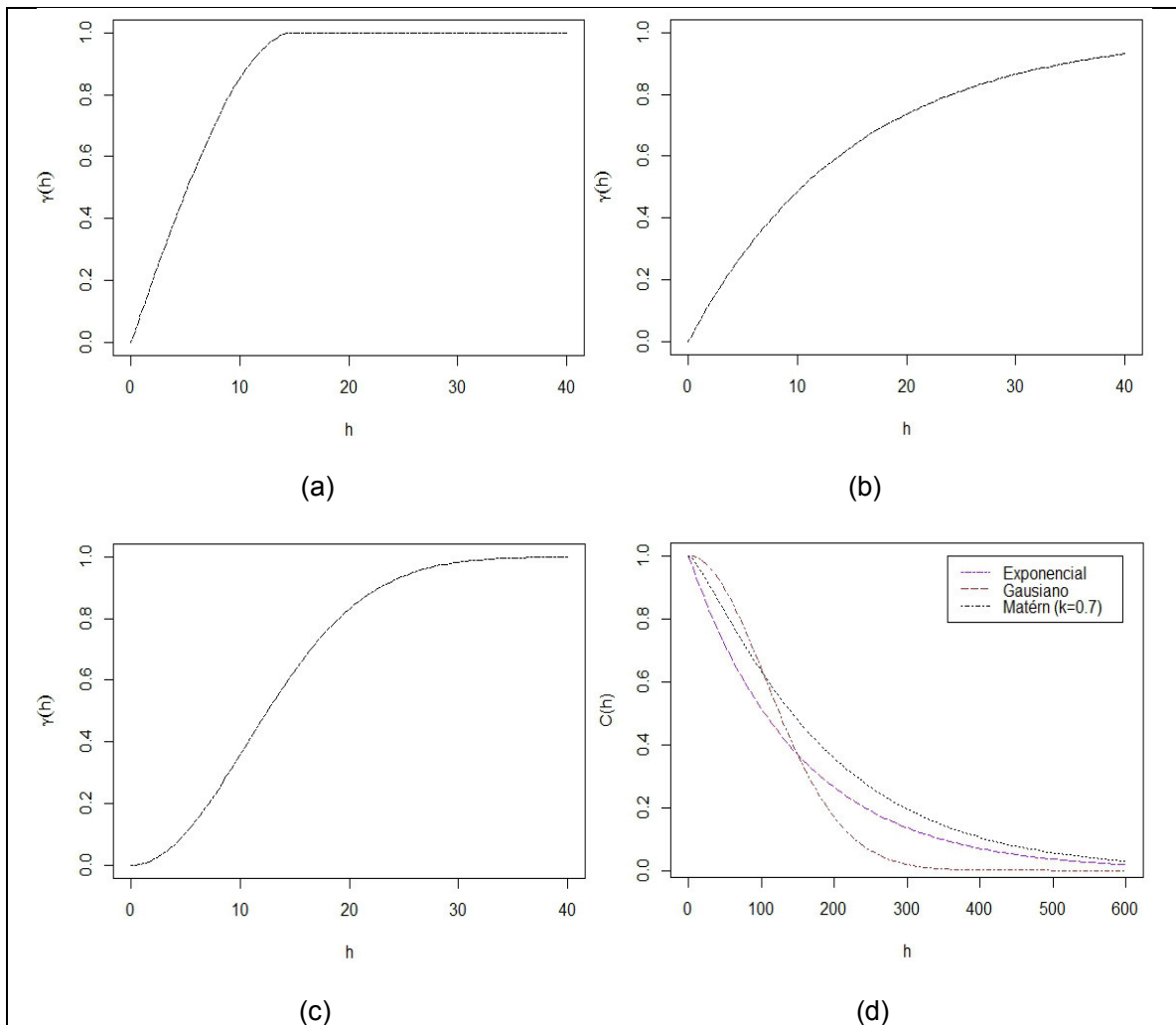


Figura 4. Gráficos de los modelos esférico (a), exponencial (b), gaussiano (c) para la función semivarianza $\gamma(h)$ y gráficos de los modelos exponencial, gaussiano y de la familia Matérn con parámetro de orden 0.7 para las función de covarianza $C(h)$ (d).

1.5 Estimación de parámetros en el ajuste de modelos teóricos a semivariogramas

Seleccionar entre los modelos espaciales teóricos uno que represente adecuadamente las función semivarianza muestral, no es una tarea fácil. Es necesario conocer los métodos de estimación de parámetros y criterios de selección del mejor modelo ajustado. Diversos autores tales como McBRATNEY & WEBSTER (1986), ISAKS & SRIVASTAVA, (1989), CRESSIE (1993), DIGGLE & RIBEIRO JR. (2007) entre otros, destacan la importancia de ajustar un modelo teórico adecuado al semivariograma experimental.

Seleccionar un buen modelo que ajuste a los datos no es un procedimiento automático, requiere un buen juicio, experiencia y comprensión de las limitaciones de los modelos teóricos, según McBRATNEY & WEBSTER (1986).

Los métodos estadísticos más frecuentes de estimación de los parámetros $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \varphi_3)^T$ que definen la estructura de dependencia espacial presentados en la literatura son: Mínimos Cuadrados Ordinarios, Mínimos Cuadrados Generalizados, Mínimos Cuadrados Ponderados, Máxima Verosimilitud y Máxima Verosimilitud Restringida.

1.5.1 Método de mínimos cuadrados ordinarios (OLS)

Sea $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \varphi_3)^T$ el vector de parámetros desconocidos de un determinado modelo espacial. El método de mínimos cuadrados ordinarios consiste en minimizar la función:

$$\sum_{j=1}^k (\hat{\gamma}(h_j) - \gamma(h_j, \boldsymbol{\varphi}))^2, \quad (20)$$

donde, $\hat{\gamma}(h_j)$ es el valor estimado de la función semivarianza experimental correspondiente a la distancia h_j , utilizando el estimador de Matheron (ecuación 7) y $\gamma(h_j, \boldsymbol{\varphi})$ es la semivarianza correspondiente a la distancia h_j dada por el modelo espacial teórico seleccionado para ser ajustado, evaluado en $\boldsymbol{\varphi} = \hat{\boldsymbol{\varphi}}$.

Diversos autores critican el método de los mínimos cuadrados ordinarios. McBRATNEY & WEBSTER (1986) destacan que este procedimiento asume que los residuos son

normalmente distribuidos e independientes, y que las funciones semivarianza $\hat{\gamma}(h_j)$ tiene todas las mismas varianzas, siendo inadecuado en el ajuste de semivariogramas. Según CRESSIE (1985), el método de mínimos cuadrados ordinarios es apenas un criterio numérico.

URIBE-OPAZO & KAVANAGH (2006) constataron que la calidad del ajuste puede ser mejorada con el uso de mínimos cuadrados ordinarios unidos a las restricciones en relación a los valores de los parámetros, restricciones que pueden ser impuestas a partir del conocimiento de las características de la variable en estudio.

1.5.2 Método de mínimos cuadrados generalizados (GLS)

Según CRESSIE (1985) la estimación de parámetros utilizando el método de mínimos cuadrados generalizados, busca superar las deficiencias del método de los mínimos cuadrados ordinarios.

Sea $\hat{\gamma}(\mathbf{h}) = (\hat{\gamma}(h_1), \dots, \hat{\gamma}(h_k))^T$ un vector $k \times 1$ de variables aleatorias con matriz de covarianza \mathbf{V} . De acuerdo con el método de los mínimos cuadrados generalizados, obtendremos el vector de parámetros que minimiza la la función:

$$(\hat{\gamma}(\mathbf{h}) - \gamma(\mathbf{h}, \boldsymbol{\varphi}))^T \mathbf{V}^{-1} (\hat{\gamma}(\mathbf{h}) - \gamma(\mathbf{h}, \boldsymbol{\varphi})), \quad (21)$$

donde, $\gamma(\mathbf{h}, \boldsymbol{\varphi})$ es el vector $k \times 1$ de semivarianzas del modelo teórico ajustado evaluado en $\boldsymbol{\varphi} = \hat{\boldsymbol{\varphi}}$ para las distancias, h_1, \dots, h_k .

CRESSIE (1993) hace mención a la dificultad de calcular la matriz de covarianza \mathbf{V} . McBRATNEY & WEBSTER (1986) destacan la dificultad de minimizar la función (21).

1.5.3 Método mínimos cuadrados ponderados (WLS)

El método de mínimos cuadrados ponderados consiste en minimizar la función:

$$\sum_{j=1}^k w_j (\hat{\gamma}(h_j) - \gamma(h_j, \boldsymbol{\varphi}))^2, \quad (22)$$

donde, los pesos w_j , atribuidos a los k puntos de la función semivarianza experimental, son seleccionados de acuerdo con um criterio estadístico.

En la literatura actual, tal como refieren CRESSIE (1993); DIGGLE & RIBEIRO JR. (2007) y en los aplicativos computacionales, tal como el IDRISI 32 (EASTMAN, 2001) disponibiliza a los usuarios tres criterios de ajuste por el método de los mínimos cuadrados ponderados, llamados de WLS1, WLS2 y WLS3.

En el método de estimación WLS1, los pesos atribuidos a los diversos puntos de la función semivarianza experimental son directamente proporcionales al número de pares de puntos de la muestra correspondientes a cada semivarianza estimada. Es considerado como el mejor vector de parámetros $\boldsymbol{\varphi} = \hat{\boldsymbol{\varphi}}$ aquel que minimiza la función:

$$\sum_{j=1}^k N(h_j) (\hat{\gamma}(h_j) - \gamma(h_j, \boldsymbol{\varphi}))^2. \quad (23)$$

En el método de estimación WLS2, los pesos atribuidos son proporcionales al número de pares de puntos de la muestra que contribuyen para la estimación de la semivarianza en cada distancia e inversamente proporcional al cuadrado de la semivarianza estimada. La expresión a ser minimizada es la ecuación (24) para $\boldsymbol{\varphi} = \hat{\boldsymbol{\varphi}}$:

$$\sum_{j=1}^k \frac{N(h_j)}{[\hat{\gamma}(h_j)]^2} (\hat{\gamma}(h_j) - \gamma(h_j, \boldsymbol{\varphi}))^2. \quad (24)$$

CRESSIE (1985) propone que los pesos w_j sean inversamente proporcionales a la $\text{Var}[\gamma(h)]$ y demuestra que, para procesos gaussianos y semivarianzas estimadas por el estimador clásico de Matheron, se tiene que: $\text{Var}[\gamma(h)] \approx \frac{[\gamma(h, \boldsymbol{\varphi})]^2}{N(h_j)}$.

De esta manera, CRESSIE (1993) propone el método de estimación WLS3 en la que la expresión a ser minimizada es la función:

$$\sum_{j=1}^k \frac{N(h_j)}{[\gamma(h_j, \boldsymbol{\varphi})]^2} \left(\hat{\gamma}(h_j) - \gamma(h_j, \boldsymbol{\varphi}) \right)^2. \quad (25)$$

Se observa, que en la ecuación (25), los pesos atribuidos a cada punto de la función semivarianza no son fijos, por ser función de $\boldsymbol{\varphi}$.

PANNATIER (1996) presenta en el software VARIOWIN 2.2 un método de estimación también considerado mínimos cuadrados, que no ofrece un ajuste automático, sino que posibilita al usuario un ajuste visual acompañado de un índice de adecuación del ajuste, calculado con base en el criterio de mínimos cuadrados ponderados. El peso atribuido a cada punto de la función semivarianza experimental, en el cálculo del índice de adecuación, es directamente proporcional al número de pares correspondiente e inversamente proporcional al valor de h .

En la búsqueda de la convergencia para el mejor ajuste, PANNATIER (1996) usa el Índice de Mejor Ajuste – IGF, valor adimensional, que minimiza la función:

$$IGF = \frac{1}{N} \sum_{k=1}^N \sum_{i=0}^{n(k)} \frac{P(i)}{\sum_{j=0}^{n(k)} P(j)} \cdot \frac{D(k)}{d(i)} \cdot \left[\frac{\gamma(h_i, \boldsymbol{\varphi}) - \hat{\gamma}(h_i)}{\sigma^2} \right]^2, \quad (26)$$

donde, N : número de variogramas direccionales; $n(k)$: número de distancias relativa al k -ésimo variograma; $D(k)$: distancia máxima en el k -ésimo variograma; $P(i)$: número de pares del i -ésima distancia del k -ésimo variograma; $d(i)$: distancia media de los pares de la i -ésima distancia del k -ésimo variograma; σ^2 : varianza de los datos.

En cualquier de los casos, WLS1, WLS2, WLS3 y IGF, los ajustes son realizados de forma iterativa, a partir de un vector de parámetros inicial $\boldsymbol{\varphi}_0$, que puede ser seleccionado por medio de un ajuste visual.

DIGGLE & RIBEIRO JR. (2007) observan que, sin el uso de restricciones adecuadas con respecto a los parámetros, a ser impuestas en el problema de optimización, el método de mínimos cuadrados puede presentar como “mejor” ajuste modelos no adecuados, estimando como por ejemplo, parámetros negativos. Además, los autores destacan que los parámetros

estimados pueden ser afectados por el valor del punto de corte (*cutoff*), que determina la distancia del campo geométrico para la cual la semivarianza es estimada.

1.5.4 Método de máxima verosimilitud (ML)

El método de estimación de máxima verosimilitud (ML) es muy utilizado en la teoría estadística. Para la estimación del vector de parámetros desconocidos $\theta = (\mu, \varphi^T)^T$ del modelo lineal espacial definido en las ecuaciones (1) y (2), se selecciona un vector que maximiza la función de verosimilitud en el dominio \mathcal{O} , siendo \mathcal{O} el espacio paramétrico.

Sea $\{Z(s), s \in \mathbf{S}\}$, $\mathbf{S} \subset \mathbb{R}^2$, un proceso gaussiano, esto es, \mathbf{Z} tiene distribución normal n -variada con vector de medias $\mu\mathbf{1}$ y matriz de covarianza $\Sigma = \Sigma(\varphi) = \varphi_1 \mathbf{I}_n + \varphi_2 \mathbf{R}(\varphi_3)$, que se denota por $\mathbf{Z} \sim N_n(\mu\mathbf{1}, \Sigma)$, donde, μ es una constante, $\mathbf{1}$ es un vector de unos, $n \times 1$ y Σ es una matriz simétrica, definida positiva $n \times n$, dada en la ecuación (2).

El método de estimación de máxima verosimilitud consiste en maximizar la función de verosimilitud o el logaritmo de la función de verosimilitud presentado:

$$l(\theta) = -\frac{n}{2} \log 2\pi - \frac{1}{2} \log |\Sigma| - \frac{1}{2} (\mathbf{Z} - \mu\mathbf{1})^T \Sigma^{-1} (\mathbf{Z} - \mu\mathbf{1}). \quad (27)$$

El logaritmo de la función de verosimilitud es básicamente una función del vector de parámetros $\theta = (\mu, \varphi^T)^T$, entonces la mejor estimación para el vector de parámetros será aquella que maximiza el logaritmo de la función de verosimilitud, esto es, $l(\hat{\theta}) = \max l(\theta)$, $\theta = (\mu, \varphi^T)^T$. En la Sección 4 se presenta el método de máxima verosimilitud para modelos espaciales lineales.

1.5.5 Método de máxima verosimilitud restringida (MLR)

Otro método utilizado para estimar los parámetros de la función de covarianza $\Sigma = \Sigma(\varphi)$ es el de Máxima Verosimilitud Restringida – MLR (ZIMMERMAN, 1989; CRESSIE, 1993; CHRISTENSEN et al., 1993).

Según SCHABENBERGER & GOTWAY (2005), un estimador de máxima verosimilitud restringido de $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \varphi_3)^T$ es definido como el vector $\hat{\boldsymbol{\varphi}}$ perteneciente al dominio de Θ , que maximiza el logaritmo de la función de verosimilitud de $\mathbf{W} = \mathbf{CZ}$, donde \mathbf{C} , $(n-1) \times n$, es la matriz de contrastes linealmente independientes (un contraste es cualquier combinación lineal $\mathbf{W} = \mathbf{CZ}$, tal que, $E[\mathbf{CZ}] = \mathbf{0}$).

Si $\mathbf{Z} \sim N_n(\mu \mathbf{1}, \boldsymbol{\Sigma})$, entonces, $\mathbf{W} = \mathbf{CZ} \sim N_{n-1}(\mathbf{0}, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T)$ y el estimador de máxima verosimilitud restringido de $\boldsymbol{\varphi}$ consiste en maximizar el logaritmo de la función de verosimilitud restringida, definida como:

$$l_R(\boldsymbol{\varphi}) = -\frac{(n-1)}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T| - \frac{1}{2} \mathbf{W}^T (\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^T)^{-1} \mathbf{W}.$$

(28)

Luego, la mejor estimación para el vector de parámetros $\boldsymbol{\varphi}$, es aquella que maximiza la función de verosimilitud restringida o el logaritmo de la función verosimilitud restringida, esto es, $l_R(\hat{\boldsymbol{\varphi}}) = \max l_R(\boldsymbol{\varphi})$.

1.6 Kriging

El kriging es una técnica de interpolación geoestadística que considera las características de autocorrelación espacial de las variables regionalizadas, en las que debe existir cierta continuidad espacial, lo que permite que los datos obtenidos mediante el muestreo de determinados puntos pueden ser utilizados para parametrizar la estimación de los puntos cuya valor de la variable es desconocido.

El termino kriging es derivado del nombre Daniel G. Krige (KRIGE, 1951), que fue el pionero en la introducción del uso de medias móviles para evitar la sobre estimación sistemática de reservas de minerales. Inicialmente el método de kriging fue desarrollado para solucionar problemas de mapeamiento geológicos, pero su uso se multiplicó con éxitos en el

mapeamiento de suelos (BURGESS & WEBSTER, 1980), mapeamientos hidrológicos (KITANIDIS & VOMVORIS, 1983) y en otras áreas.

La diferencia entre el kriging y otros métodos de interpolación es la manera como los pesos son atribuidos a los diferentes elementos de la muestra. En el caso de interpolación lineal simple, por ejemplo, los pesos son todos iguales a $1/n$ (n es el número de elementos de la muestra); en la interpolación basada en el inverso del cuadrado de las distancias, los pesos son definidos como la inversa del cuadrado de la distancia que separa el valor interpolado de los valores observados. En la interpolación por kriging, el procedimiento es semejante al de la interpolación por medias móviles ponderadas, excepto que aquí los pesos son determinados a partir de un análisis espacial, basado en el estudio de la estructura de variabilidad espacial de los datos espacialmente georreferenciados. El kriging ofrece estimaciones insesgadas y de varianzas mínimas (WEBSTER & OLIVER, 2007).

1.6.1 Kriging ordinario

Según ISAACS & SRIVASTAVA (1989) y JOURNEL (1989) un valor estimado Z en el punto s_0 , $\hat{Z}(s_0)$, puede ser obtenida a partir de una serie de valores medidos $Z(\mathbf{s}_i)$, por medio de una combinación lineal ponderada, asociando un peso λ_i , $i = 1, \dots, n$ para cada valor medido. Con respecto al número de observaciones a ser usadas en esta estimación, los autores resaltan que si el fenómeno presenta continuidad, entonces se puede usar los puntos vecinos más próximos. En kriging, esos n valores conocidos proceden de una misma variable aleatoria regionalizada. El estimador puede ser escrito por la ecuación:

$$\hat{Z}(s_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i), \quad (29)$$

donde, $\hat{Z}(s_0)$: es valor estimado en el punto s_0 , no muestreado; n : número de valores medidos $Z(\mathbf{s}_i)$ de las variables en los puntos muestreados; λ_i : es un peso asociado a cada valor $Z(\mathbf{s}_i)$ medido.

El estimador $\hat{Z}(s_0)$ definido en la ecuación (29) es considerado como el mejor estimador insesgado de varianza mínima, esta característica se debe a que los pesos λ_i , son obtenidos suponiendo que la esperanza del error es cero y la varianza del error es mínima (WEBSTER & OLIVER, 2007).

Sobre la hipótesis que el proceso estocástico \mathbf{Z} es estacionario, se tiene que el error de estimación $\varepsilon(s_0) = Z(s_0) - \hat{Z}(s_0)$, tiene esperanza:

$$E[\varepsilon(s_0)] = E[Z(s_0) - \hat{Z}(s_0)] = E[Z(s_0) - \sum_{i=1}^n \lambda_i Z(s_i)] = \mu - \mu \sum_{i=1}^n \lambda_i = \mu (1 - \sum_{i=1}^n \lambda_i),$$

el error de estimación será nulo si:

$$\sum_{i=1}^n \lambda_i = 1. \quad (30)$$

La condición definida en la ecuación (30) es necesaria para garantizar que el estimador no sea sesgado.

Sea la varianza σ^2_E del error de estimación $\varepsilon(s_0) = Z(s_0) - \hat{Z}(s_0)$;

$$\sigma^2_E = \text{Var}[\varepsilon(s_0)] = \text{Var}[Z(s_0) - \hat{Z}(s_0)] = E[(Z(s_0) - \hat{Z}(s_0))^2] - \{E[Z(s_0) - \hat{Z}(s_0)]\}^2,$$

$$\sigma^2_E = E[(Z(s_0) - \hat{Z}(s_0))^2] = E[(Z(s_0) - \sum_{i=1}^n \lambda_i Z(s_i))^2],$$

descomponiendo el termino cuadrático,

$$\sigma^2_E = E[(Z(s_0))^2] + E[\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j Z(s_i) Z(s_j)] - 2 E[\sum_{i=1}^n \lambda_i Z(s_i) Z(s_0)]$$

$$\sigma^2_E = \text{Var}[Z(s_0)] + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \text{Cov}[Z(s_i), Z(s_j)] - 2 \sum_{i=1}^n \lambda_i \text{Cov}[Z(s_i), Z(s_0)], \quad (31)$$

que es equivalente a resolver

$$\sigma^2_E = \text{Var}[Z(s_0)] + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \text{Cov}[Z(s_i), Z(s_j)] - 2 \sum_{i=1}^n \lambda_i \text{Cov}[Z(s_i), Z(s_0)] - 2\eta (1 - \sum_{i=1}^n \lambda_i),$$

introduciendo el parámetro η de lagrange a la ecuacion (31), se tiene por la ecuación (30)

$$\text{que: } 2\eta (1 - \sum_{i=1}^n \lambda_i) = 0.$$

La minimización de la varianza σ^2_E consiste, en calcular las $n+1$ derivadas parciales con relación a $\lambda_1, \dots, \lambda_n, \eta$ e igualar a cero, obteniendo un sistema de $n+1$ ecuaciones y $n+1$ incógnitas cuya solución resulta en los n pesos λ_i que cumplen la condición de ser insesgado

$\sum_{i=1}^n \lambda_i = 1$ y, al mismo tiempo minimiza la varianza de estimación

$$\text{Var}[Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0)] - 2\eta(1 - \sum_{i=1}^n \lambda_i).$$

El desarrollo de las $n+1$ derivadas en relación a $\lambda_1, \dots, \lambda_n, \eta$, se tiene que:

$$2 \sum_{j=1}^n \lambda_j \text{Cov}[Z(\mathbf{s}_i), Z(\mathbf{s}_j)] - 2\text{Cov}[Z(\mathbf{s}_i), Z(\mathbf{s}_0)] - 2\eta = 0, \text{ para } i=1, \dots, n, \text{ y}$$

$$2(1 - \sum_{i=1}^n \lambda_i) = 0.$$

El sistema kriging de $n+1$ ecuaciones, de cuya solución obtenemos los n pesos λ_i que son dados de la solución de:

$$\sum_{j=1}^n \lambda_j \text{Cov}[Z(\mathbf{s}_i), Z(\mathbf{s}_j)] - \eta = \text{Cov}[Z(\mathbf{s}_i), Z(\mathbf{s}_0)], \text{ para } i=1, \dots, n; \sum_{i=1}^n \lambda_i = 1.$$

$$\text{Luego, } \sigma^2_E = \text{Var}[Z(\mathbf{s}_0)] - \sum_{i=1}^n \lambda_i \text{Cov}[Z(\mathbf{s}_i), Z(\mathbf{s}_0)] - \eta.$$

De la relación, $\gamma(h) = C(0) - C(h)$, se tiene que las condiciones que garantizan insesgamiento y la varianza mínima, son soluciones del sistema de ecuaciones dadas en (32) para $h_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$ que es la distancia euclidiana entre \mathbf{s}_i y \mathbf{s}_j :

$$\sum_{j=1}^n \lambda_j \gamma(h_{ij}) + \eta = \gamma(h_{0i}), \text{ para } i=1, \dots, n; \sum_{i=1}^n \lambda_i = 1. \quad (32)$$

La solución de la ecuación (32) garantizan que el estimador de kriging es el mejor estimador lineal insesgado de varianza mínima “BLUE” (*Best Linear Unbiased Estimator*), (WEBSTER & OLIVER, 2007).

El sistema de ecuaciones en (32), es constituido de $n+1$ ecuaciones y $n+1$ incógnitas, es conocido como sistema de kriging ordinaria. La incógnita η es un multiplicador de Lagrange,

introducido al minimizar la varianza del error. En forma matricial, el sistema puede ser escrito como: $\gamma\lambda = \mathbf{A}$,

donde, $\lambda = [\lambda_1 \ \lambda_2 \ \lambda_3 \ \dots \ \lambda_n \ \eta]^T$, vector $n+1 \times 1$;

$$\gamma = \begin{bmatrix} \gamma(h_{11}) & \gamma(h_{12}) & \dots & \gamma(h_{1n}) & 1 \\ \gamma(h_{21}) & \gamma(h_{22}) & \dots & \gamma(h_{2n}) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(h_{n1}) & \gamma(h_{n2}) & \dots & \gamma(h_{nn}) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix}, \text{ es una matriz } n+1 \times n+1;$$

$\mathbf{A} = (\gamma(h_{10}), \dots, \gamma(h_{n0}), 1)^T$, es un vector $n+1 \times 1$.

Los pesos λ_i pueden, ser obtenidos mediante el producto de la matriz inversa de γ por la matriz \mathbf{A} , dado como:

$$\lambda = \gamma^{-1} \mathbf{A}. \quad (33)$$

Una vez obtenidos los pesos λ_i y el valor η , se puede también calcular la varianza del error σ^2_E definida en la ecuación (31) que es dada por CRESSIE, (1993) como:

$$\sigma^2_E = \sum_{i=1}^n \lambda_i \gamma(h_{i0}) - \eta. \quad (34)$$

Se observa en la ecuación (34) que los elementos de la matriz \mathbf{A} requieren información del valor de la función semivarianza en h_{i0} , que no necesariamente son iguales a las distancias del semivariograma experimental, de esta forma es necesario el ajuste de un modelo teórico al semivariograma experimental.

1.6.2 Kriging universal

Al modelar la variabilidad espacial existen problemas tales como la existencia de tendencia espacial de los datos, esto es, los valores medidos aumentan o disminuyen en alguna dirección en el área en estudio. Este es el caso de fenómenos no estacionarios, que hacen imposible la aplicación del kriging ordinaria (KRAJEWSKI & GIBBS, 1993). Con el objetivo de

dar solución a este problema, MATHERON (1970) propone la aproximación por kriging universal (KU), que consiste en extraer de la variable regionalizada $Z(\mathbf{s})$ la parte no estacionaria por medio de una componente determinística $\mu(\mathbf{s})$ que representa la tendencia, hasta encontrar la parte estacionaria del fenómeno, de esta forma se obtiene una componente estocástica $\epsilon(\mathbf{s})$ relacionada a $Z(\mathbf{s})$ por la expresión: $Z(\mathbf{s}) = \mu(\mathbf{s}) + \epsilon(\mathbf{s})$.

Para la componente determinística $\mu(\mathbf{s})$ se sugiere modelar la tendencia en la posición \mathbf{s} , de la forma dada en (35):

$$\mu(\mathbf{s}) = \sum_{l=0}^k a_l f_l(\mathbf{s}), \quad k = 0, \dots, K, \quad (35)$$

donde, a_l son los coeficientes desconocidos asociados a $f_l(\mathbf{s})$ que es una función conocida de \mathbf{s} , que describe la tendencia en la posición \mathbf{s} (ISAAKS & SRIVASTAVA, 1989; CRESSIE, 1993). Cuando no existe tendencia el kriging universal dado en la ecuación (35) se reduce a un kriging ordinaria (CHRISTENSEN et al., 1993).

Finalmente, se obtiene un sistema de ecuaciones para el kriging universal de la forma:

$$\sum_{j=1}^n \lambda_j \gamma(\mathbf{s}_i, \mathbf{s}_j) + \sum_{l=0}^k a_l f_l(\mathbf{s}_j) = \gamma(\mathbf{s}_i, \mathbf{s}_0) \quad ; \quad \sum_{i=1}^n \lambda_i f_l(\mathbf{s}_i) = f_l(\mathbf{s}_0), \quad l = 0, \dots, K,$$

con varianza de estimación dada por $\sigma_{KU}^2 = \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_i, \mathbf{s}_j) + \sum_{l=0}^k a_l f_l(\mathbf{s}_j)$, de esta forma, el valor estimado por kriging universal será $\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i)$.

1.7 Criterios de selección de modelos

1.7.1 Validación cruzada

La validación cruzada es una técnica para evaluar los errores de estimación que compara los valores observados con estimados por el modelo espacial. El procedimiento de validación cruzada es el siguiente: el valor de un elemento de la muestra en una cierta ubicación es temporalmente eliminado del conjunto de datos de la muestra, un nuevo valor en la misma

ubicación se estima por kriging, utilizando los elementos restantes de la muestra. Una vez que se calcula la estimación, se puede comparar con el valor de la muestra que se retiró inicialmente del conjunto de datos. Este procedimiento se repite para todas los elementos de la muestra disponibles y se denomina método de "dejar uno fuera" (ISAAKS & SRIVASTAVA, 1989).

A partir del conocimiento de los valores de la muestra y los valores estimados, se puede conocer la varianza total de la estimación, para evaluar la precisión del proceso. Se espera que los errores de estimación obtenidos en la ecuación (36) tengan media cero, varianza constante y distribución normal.

$$\varepsilon_i = \varepsilon(s_i) = Z(s_i) - \hat{Z}(s_{(i)}), \quad \text{para, } i = 1, \dots, n, \quad (36)$$

donde, $\hat{Z}(s_{(i)})$ es el valor estimado por kriging sin la i -ésima observación $Z(s_i)$.

El error de estimación no indica solamente la eficiencia del ajuste de los diferentes modelos teóricos al semivariograma experimental y el modelamiento del proceso, sino que también indica la evaluación de la estacionariedad y la influencia de datos atípicos o discrepantes ("outliers"). La comparación puede ser efectuada por medio del error medio (EM), error medio reducido (\overline{ER}), desviación estándar de los errores $\varepsilon_1, \dots, \varepsilon_n$ (S_{EM}), desviación estándar de los errores reducidos ER_1, \dots, ER_n (S_{ER}) y del error absoluto (EA).

El error medio es definido como:

$$EM = \frac{1}{n} \sum_{i=1}^n \varepsilon_i = \frac{1}{n} \sum_{i=1}^n \left(Z(s_i) - \hat{Z}(s_{(i)}) \right). \quad (37)$$

donde, n es el número de datos; $Z(s_i)$ es valor observado en el punto s_i ; $\hat{Z}(s_{(i)})$ y el valor estimado por kriging en el punto s_i sin considerar la observación $Z(s_i)$.

El error medio reducido es definido por MCBRATNEY & WEBSTER (1986) y CRESSIE (1993), como:

$$\overline{ER} = \frac{1}{n} \sum_{i=1}^n ER_i = \frac{1}{n} \sum_{i=1}^n \left(\frac{Z(s_i) - \hat{Z}(s_{(i)})}{\sigma(\hat{Z}(s_{(i)}))} \right), \quad (38)$$

donde, $\sigma(\hat{Z}(s_{(i)}))$ es la desviación estándar del valor estimado por kriging en el punto s_i sin considerar la observación $Z(s_i)$.

Según MCBRATNEY & WEBSTER (1986); CRESSIE (1993); VIEIRA (2000); MELLO et al. (2005) y FARACO et al. (2008) en condiciones de no insesgado, el valor para el error medio reducido debe ser cero y desviación estándar del error medio reducido debe ser igual a uno. Por tanto, los valores del error medio EM y error medio reducido \overline{ER} que esten próximos de cero, el menor valor de la desviación estándar del error medio (S_{EM}) y el valor de la desviación estándar del error medio reducido (S_{ER}) esten próximo de uno, son los criterios para elegir el mejor modelo ajustado.

Otra medida de selección de modelos es el error absoluto (EA) que es una medida de magnitud de los errores en la misma unidad de medida de la variable en estudio. El EA es definido como:

$$EA = \sum_{i=1}^n |Z(s_i) - \hat{Z}(s_{(i)})|, \quad (39)$$

donde se procura el menor valor del error absoluto.

1.7.2 Criterios de información de Akaike (AIC)

El criterio de Información de Akaike - AIC (AKAIKE, 1973) trata de demostrar que si dos modelos representan datos igualmente satisfactorios, del modelo más simple se puede esperar un mejor desempeño para la predicción de nuevos datos. Por lo tanto, este criterio busca la simplicidad, imponiendo una sanción por la complejidad. El AIC es definido como:

$$AIC = -2l(\theta) + 2u, \quad (40)$$

donde, $l(\theta)$ es el logaritmo de la función verosimilitud evaluado en $\theta = \hat{\theta}$; u es el número de parámetros del modelo ajustado $u = p + 3$.

La decisión de elegir entre los modelos utilizados en el ajuste, recae sobre aquel modelo que presenta el menor valor de AIC .

Un problema con el criterio AIC es que se basa en el logaritmo de la función verosimilitud y, por tanto, el número n de observaciones no puede ser pequeña ($n < 40$), donde s es la desviación estándar de los datos. En estos casos, es preferible utilizar el criterio de

información de Akaike de segunda orden (*CAIC*), presentado como: $CAIC = -2l(\theta) + 2u + \frac{2u(u+1)}{n-u-2}$.

El valor de *CAIC* tiende para *AIC*, cuando n crece. La decisión para seleccionar entre los modelos utilizados en el ajuste recae sobre aquel modelo que presenta el menor valor de *CAIC*.

1.7.3 Criterios de información de Schwarz (*SIC*)

El criterio de información de Schwarz-*SIC* (SCHWARZ, 1978) es definido como:

$$SIC = -2l(\theta) + u \log(n), \quad (41)$$

donde, $l(\theta)$ es el logaritmo de la función verosimilitud evaluada en $\theta = \hat{\theta}$; u es el número de parámetros del modelo ajustado $u = p + 3$ y n es el número de datos.

CAPÍTULO 2

COMPARACIÓN DE MAPAS TEMÁTICOS

Un procedimiento utilizado para la comparación de dos mapas temáticos en la cual uno de ellos es denominado mapa de referencia y el otro mapa modelo es el cálculo de los índices de exactitud obtenidos a partir de una matriz denominada matriz de errores.

La Tabla 1 presenta en forma general la matriz de los errores. En esta matriz, los *pixeles*¹ del mapa de referencia son cuantificados en las columnas, en cuanto que los *pixeles* del mapa modelo son cuantificados en las filas. Cada elemento de la matriz representa la cantidad de *pixeles* pertenecientes a la clase C_i , $i = 1, \dots, k$, del mapa modelo y a la clase C_j , $j = 1, \dots, k$, del mapa de referencia. Es importante destacar que para comparar los mapas temáticos se debe tener las mismas clases en ambos mapas.

Tabla 1. Matriz de errores de orden $k \times k$.

| Mapa de referencia | | | | | | |
|-------------------------------|----------|----------|----------|----------|----------|----------------------------|
| Mapa modelo | Clases | C_1 | C_2 | \dots | C_k | Total por fila n_{i+} |
| | C_1 | n_{11} | n_{12} | \dots | n_{1k} | n_{1+} |
| | C_2 | n_{21} | n_{22} | \dots | n_{2k} | n_{2+} |
| | \vdots | \vdots | | \vdots | \vdots | \vdots |
| | \cdot | \cdot | | \cdot | \cdot | \cdot |
| | C_k | n_{k1} | n_{k2} | \dots | n_{kk} | n_{k+} |
| Total por columna n_{+j} | | n_{+1} | n_{+2} | \dots | n_{+k} | Total general N |

k : número de clases; C_i : clase i ; n_{i+} : total de *pixeles* en la clase C_i del mapa modelo; n_{+j} : total de *pixeles* en la clase C_j del mapa de referencia; N : número total de *pixeles*.

Sea N el número total de *pixeles*, los elementos de la diagonal principal n_{ii} ($i = j$) representan la situación en que los *pixeles* tuvieron la misma clasificación en ambos mapas, en cuanto que los elementos fuera de la diagonal principal representan las clasificaciones

¹ Aglutinación de las palabras *Picture* y *Elements*. Representa el menor elemento de un mapa al cual es posible atribuir un valor. En este trabajo los *pixeles* serán definidos como los elementos de la matriz de los valores kriging.

erradas, siendo que $n_{i+} = \sum_{j=1}^k n_{ij}$ es la cantidad de *pixeles* (área) de la fila C_i , y $n_{+i} = \sum_{u=1}^k n_{ui}$ es la cantidad de *pixeles* (área) de la columna C_i .

Utilizando los elementos de la matriz de errores es obtenido el índice de Exactitud Global (EG) (CONGALTON & GREEN, 1999; ANDERSON et al., 2001; BACH et al., 2006), El índice Kappa (K) (CONGALTON & GREEN, 1999; JENNESS & WYNNE, 2005; DE BASTIANI et al., 2012); y el índice Tau (T) (MA & REDMOND, 1995), presentados respectivamente en las ecuaciones (42), (43) y (44).

$$EG = \frac{\sum_{i=1}^k n_{ii}}{N} ; \quad (42)$$

$$K = \frac{N \sum_{i=1}^k n_{ii} - \sum_{i=1}^k n_{i+} n_{+i}}{N^2 - \sum_{i=1}^k n_{i+} n_{+i}} ; \quad (43)$$

$$T = \frac{\frac{\sum_{i=1}^k n_{ii}}{N} - p_k}{1 - p_k} . \quad (44)$$

El índice de exactitud global (EG) dado en la ecuación (42) es una medida de evaluación utilizada para medir la similaridad entre el mapa de referencia y el mapa modelo, que según ANDERSON *et al.* (2001) el nivel mínimo de similaridad es de $EG \geq 0.85$. Es posible construir un intervalo de $(1-\alpha)\%$ de confianza para EG , siendo que este es de la forma $IC[EG, (1-\alpha)\%] = EG \pm z_{\alpha/2} \sqrt{EG(1-EG)/N}$, para $N > 50$ donde, $Z_{\alpha/2}$ es el valor nominal de una variable aleatoria con distribución normal estandarizada (CONGALTON & GREEN, 1999).

El índice Kappa (K) (CONGALTON & GREEN, 1999; DE BASTIANI et al., 2012) dado en la ecuación (43) viene siendo utilizado para medir la exactitud de las clasificaciones temáticas, siendo recomendado como una medida apropiada de la exactitud por utilizar todos los elementos de la matriz de errores. El índice Kappa (K) nos ofrece una medida de concordancia entre los valores del mapa de referencia y los valores del mapa modelo. Según

la clasificación de KRIPENDORFF (2004), K es clasificado como baja exactitud si $K < 0.67$, media exactitud si $0.67 \leq K < 0.80$ y alta exactitud si $K \geq 0.80$. La varianza del índice Kappa es obtenida como,

$$\sigma^2(K) = \frac{1}{N} \left(\frac{\theta_1(1-\theta_1)}{(1-\theta_2)^2} + \frac{2(1-\theta_1)(2\theta_1\theta_2-\theta_3)}{(1-\theta_2)^3} + \frac{(1-\theta_1)^2(\theta_4-4\theta_2)^2}{(1-\theta_2)^4} \right), \quad (45)$$

donde, $\theta_1 = \frac{1}{N} \sum_{i=1}^k n_{ii}$; $\theta_2 = \frac{1}{N^2} \sum_{i=1}^k n_{i+} n_{+i}$; $\theta_3 = \frac{1}{N^2} \sum_{i=1}^k n_{ii} (n_{i+} + n_{+i})$; $\theta_4 = \frac{1}{N^3} \sum_{i=1}^k \sum_{j=1}^k n_{ij} (n_{i+} + n_{+j})^2$

El intervalo de $(1-\alpha)\%$ de confianza para el índice Kappa es de la forma

$$IC[K, (1-\alpha)\%] = K \pm z_{\alpha/2} \sqrt{\sigma^2(K)}, \text{ donde, } z_{\alpha/2} \text{ es el valor nominal de una variable aleatoria con distribución normal estandarizada.}$$

El índice de concordancia Tau (T) dado en la ecuación (44) (MA & REDMOND, 1995), también es conocido como índice Kappa modificado, siendo p_i la probabilidad *a priori* para cada clase i . Cuando las probabilidades *a priori* para cada clase es igual, se tiene que $p_i = 1/k$, donde k representa el número de clases de la matriz de errores. El índice Tau puede seguir la misma clasificación del Kappa. La varianza del índice Tau es calculada con la misma fórmula de la varianza del índice Kappa, la diferencia está en la incorporación de la probabilidad *a priori*.

CAPÍTULO 3

MODELO ESPACIAL LINEAR GAUSIANO

Considere un proceso estocástico gaussiano $\{Z(\mathbf{s}_i), \mathbf{s}_i \in \mathbf{S}\}$ donde $\mathbf{S} \subset \mathbb{R}^2$, definido en las ecuaciones (1) y (2). Para algunas funciones conocidas de \mathbf{s}_i , se asume que $x_1(\mathbf{s}_i), \dots, x_p(\mathbf{s}_i)$, la media del proceso estocástico, es dada como:

$$\mu(\mathbf{s}_i) = \sum_{u=1}^p x_u(\mathbf{s}_i) \beta_u, \quad i = 1, \dots, n, \quad (46)$$

donde, los parámetros β_1, \dots, β_p son desconocidos y deben ser estimados.

La función de covarianza $C(\mathbf{s}_i, \mathbf{s}_u)$ es usada en el estudio de la dependencia espacial del proceso intrínsecamente estacionario y también especificada por un vector tri-dimensional $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \varphi_3)^T$.

Por simplicidad, se puede utilizar las siguientes notaciones: $Z(\mathbf{s}_i) = z_i$, $\mathbf{Z} = (z_1, \dots, z_n)^T$, $x_{iu} = x_u(\mathbf{s}_i)$, $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ip})$, donde \mathbf{X} es una matriz, $n \times p$, de columnas con rango completo, con sus filas \mathbf{x}_i^T , $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$, $\varepsilon_i = \varepsilon(\mathbf{s}_i)$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T$, con $i = 1, \dots, n$, $u = 1, \dots, p$.

Por tanto, la media del proceso $\mu(\mathbf{s}_i) = \mathbf{x}_i^T \boldsymbol{\beta}$, entonces, se obtiene como:

$$z_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i, \quad \text{con } i = 1, \dots, n, \quad (47)$$

que puede ser escrita de la forma matricial como:

$$\mathbf{Z} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}. \quad (48)$$

Sobre la hipótesis de que los errores tienen $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \boldsymbol{\Sigma})$, se tiene que el modelo espacial lineal presentado en la ecuación (48) tiene $\mathbf{Z} \sim N_n(\mathbf{X} \boldsymbol{\beta}, \boldsymbol{\Sigma})$, donde los parámetros desconocidos del modelo $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\varphi}^T)^T$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ y $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \varphi_3)^T$ pueden ser estimados, maximizando la función de verosimilitud o el logaritmo de la función de verosimilitud.

Sea el logaritmo de la función de verosimilitud, definido en la ecuación (49):

$$l(\boldsymbol{\theta}) = -\left(\frac{n}{2}\right) \log(2\pi) - \frac{1}{2} \log|\boldsymbol{\Sigma}| - \frac{1}{2} (\mathbf{Z} - \mathbf{X} \boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Z} - \mathbf{X} \boldsymbol{\beta}), \quad (49)$$

donde, $\delta = (\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})$ es conocida como distancia de Mahalanobis y

$$\boldsymbol{\Sigma} = \varphi_1 \mathbf{I}_n + \varphi_2 \mathbf{R}(\varphi_3).$$

Básicamente $l(\boldsymbol{\theta})$ es una función del vector de parámetros $\boldsymbol{\theta} \in \Theta$, donde Θ es el espacio paramétrico. Entonces, el mejor estimador para el vector de parámetros $\boldsymbol{\theta}$ será aquel que maximiza la función de verosimilitud o el logaritmo de la función de verosimilitud, esto es, $l(\hat{\boldsymbol{\theta}}) = \max l(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta$, donde $\hat{\boldsymbol{\theta}}$ es el estimador de máxima verosimilitud de $\boldsymbol{\theta}$.

3.1 Función score

Com el objetivo de obtener estimadores de máxima verosimilitud de $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\varphi}^T)^T$ para el modelo espacial lineal gaussiano dado en (48), la función score del vector $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\varphi}^T)^T$ tiene como componentes:

$$U(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \mathbf{X}^T \boldsymbol{\Sigma}^{-1} (\mathbf{Z} - \mathbf{X}\boldsymbol{\beta}); \quad U(\boldsymbol{\varphi}) = \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\varphi}} \quad \text{que tiene como } j\text{-ésimo elemento}$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \varphi_j} = -\frac{1}{2} \text{tr} \left\{ \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \right\} + \frac{1}{2} \boldsymbol{\varepsilon}^T \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}, \quad j = 1, 2, 3, \quad \text{donde } \boldsymbol{\varepsilon} = (\mathbf{Z} - \mathbf{X}\boldsymbol{\beta}).$$

Para estimar los vectores de parámetros $\boldsymbol{\beta}$ y $\boldsymbol{\varphi}$ se resuelven el sistemas de ecuaciones : $U(\boldsymbol{\beta}) = \mathbf{0}$ y $U(\boldsymbol{\varphi}) = \mathbf{0}$, obteniendo el estimador de máxima verosimilitud para $\boldsymbol{\beta}$, dado como:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z}, \quad (50)$$

siendo que para estimar el vector de parámetro $\boldsymbol{\varphi}$, no existe una solución explícita, pues la solución constituye un sistema de ecuaciones no lineales que no presenta solución en forma cerrada, para esto se debe utilizar algoritmos de optimización numérica, como por ejemplo L-BFGS-B (BYRD et al., 1995; ZHU et al., 1997).

3.2 Matriz de información observada

La matriz de información observada (hessiana) para el modelo espacial lineal gaussiano dado en (48) es definida por $l(\theta) = -L(\theta)$ y debe ser evaluado en $\theta = \hat{\theta}$. Su forma particionada puede ser de la forma dada en la ecuación (51) con elementos definidos en las ecuaciones (52) y (53):

$$L(\theta) = \begin{pmatrix} L_{\beta\beta} & L_{\beta\varphi} \\ L_{\varphi\beta} & L_{\varphi\varphi} \end{pmatrix}, \quad (51)$$

$$\text{donde, } L_{\beta\beta} = \frac{\partial^2 l(\theta)}{\partial \beta \partial \beta^T} = -(X^T \Sigma^{-1} X); \quad (52)$$

$$L_{\beta\varphi} = L_{\varphi\beta}^T = \frac{\partial^2 l(\theta)}{\partial \beta \partial \varphi^T}, \text{ con elementos } \frac{\partial^2 l(\theta)}{\partial \beta \partial \varphi_j} = -X^T \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_j} \Sigma^{-1} \varepsilon, j = 1, 2, 3;$$

siendo, $\varepsilon = (Z - X\beta)$,

$$L_{\varphi\varphi} = \frac{\partial^2 l(\theta)}{\partial \varphi \partial \varphi^T}, \text{ con elementos}$$

$$\begin{aligned} \frac{\partial^2 l(\theta)}{\partial \varphi_i \partial \varphi_j} = & \frac{1}{2} \text{tr} \left\{ \Sigma^{-1} \left(\frac{\partial \Sigma}{\partial \varphi_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_j} - \frac{\partial^2 \Sigma}{\partial \varphi_i \partial \varphi_j} \right) \right\} \\ & + \frac{1}{2} \varepsilon^T \Sigma^{-1} \left\{ \frac{\partial^2 \Sigma}{\partial \varphi_i \partial \varphi_j} - \frac{\partial \Sigma}{\partial \varphi_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_j} - \frac{\partial \Sigma}{\partial \varphi_j} \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_i} \right\} \Sigma^{-1} \varepsilon, i, j = 1, 2, 3. \end{aligned} \quad (53)$$

3.3 Matriz de información esperada

La matriz de información esperada de Fisher para el modelo espacial lineal gaussiano definida en (48) es de la forma $K(\theta) = E[-L(\theta)]$, es bloque diagonal y es dada por la forma (LANGE et al., 1989; MITCHELL, 1989):

$$K(\theta) = \begin{pmatrix} K(\beta) & \mathbf{0} \\ \mathbf{0} & K(\varphi) \end{pmatrix},$$

donde,

$$K(\boldsymbol{\beta}) = E[-\mathbf{L}_{\boldsymbol{\beta}\boldsymbol{\beta}}] = E\left[-\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}\right] = \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X};$$

(54)

$$K(\boldsymbol{\varphi}) = E[-\mathbf{L}_{\boldsymbol{\varphi}\boldsymbol{\varphi}}] = E\left[-\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}^T}\right], \text{ que tiene como elementos}$$

$$k_{ij}(\boldsymbol{\varphi}) = \frac{1}{2} \text{tr} \left[\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \right] \quad i, j = 1, 2, 3. \quad (55)$$

Sea $K(\boldsymbol{\theta})^{-1}$ la matriz de covarianza de $\boldsymbol{\theta}$, a partir de ella se puede estimar $K(\boldsymbol{\beta})^{-1} = \mathbf{V}_{\boldsymbol{\beta}} = [(v_{jj})] = (\mathbf{X}^T \boldsymbol{\Sigma}(\hat{\boldsymbol{\varphi}})^{-1} \mathbf{X})^{-1}$, el intervalo de $(1-\alpha)\%$ de confianza para β_j , $j=1, \dots, p$, es dado por $\beta_j \pm z_{(1-\alpha/2)} \sqrt{v_{jj}}$, en que, $z_{(1-\alpha/2)}$ es el $(1-\alpha/2)$ cuantil de una distribución normal estándar.

De la misma forma para la construcción de intervalo de confianza para $\boldsymbol{\varphi}$, $K(\boldsymbol{\varphi})^{-1} = \mathbf{V}_{\boldsymbol{\varphi}} = [(u_{jj})]$, el intervalo de $(1-\alpha)\%$ de confianza para φ_j , $j=1, 2, 3$, es dado por $\varphi_j \pm z_{(1-\alpha/2)} \sqrt{u_{jj}}$, donde, $z_{(1-\alpha/2)}$ es el $(1-\alpha/2)$ cuantil de una distribución normal estándar.

CAPÍTULO 4

DIAGNÓSTICO

Un estudio de gran importancia en el análisis de diagnóstico es la detección de observaciones influyentes, esto es, puntos que ejercen un peso desproporcionado en las estimaciones del modelo y en la significancia de los parámetros. La detección de puntos influyentes, es la técnica más conocida para evaluar el impacto de retirar una observación particular en las estimaciones de parámetros en modelos de regresión. Un problema que puede ocurrir con la eliminación individual de puntos es el denominado efecto de enmascaramiento “*masking effect*”, esto es, dejar de detectar puntos conjuntamente influyentes. La distancia de COOK (1977) es la más tradicional medida para detectar puntos influyentes que fue, originalmente, desarrollada para el modelo lineal general, que rápidamente fue asimilada y extendida para diversas clases de modelos.

En el estudio de observaciones atípicas o *outliers* en variables regionalizada, CRESSIE & HARKINS (1980) presentaron un estimador de la función semivarianza denominado robusto, para el estudio de la variabilidad espacial. McBRANTNEY & WEBSTER (1986) comentan que ese estimador de la función semivarianza presentado pelos autores es limitado al realizar estudios de simulación. GENTON (1998) comenta acerca de la poca robustez del estimador de Cressie y Harkins desde el punto de vista estadístico ante la presencia de *outliers*.

4.1 Influencia global

La distancia de Cook (COOK, 1977) es una de la mas conocida herramienta para el estudio de los diagnóstico de influencia, consiste en la eliminación individual de puntos del

conjunto de datos, y la evaluación el impacto de retirar una observación en la estimación de los parámetros del modelo de regresión.

URIBE-OPAZO et al. (2012) presentan una medida análoga a la distancia de Cook con enfoque para variables regionalizadas, denominada influencia global espacial presentada en la ecuación (56). En que, $\hat{\theta}$ es el estimador de máxima verosimilitud de $\theta = (\beta^T, \varphi^T)^T$ del modelo postulado, con $\beta = (\beta_1, \dots, \beta_p)^T$, $\varphi = (\varphi_1, \varphi_2, \varphi_3)^T$, y $\hat{\theta}_{(i)}$ es el estimador de máxima verosimilitud de θ excluyendo la i-ésima observación.

$$D_{i\theta} = \frac{1}{p+3} (\hat{\theta} - \hat{\theta}_{(i)})^T K(\theta) (\hat{\theta} - \hat{\theta}_{(i)}), \quad (56)$$

Por ser $K(\theta)$ la matriz bloque diagonal tenemos que:

$$D_{i\theta} = \alpha D_{i\beta} + (1 - \alpha) D_{i\varphi}, \quad \text{para } i=1, \dots, n,$$

donde,

$$D_{i\beta} = (\hat{\beta} - \hat{\beta}_{(i)})^T K(\beta) (\hat{\beta} - \hat{\beta}_{(i)});$$

$$D_{i\varphi} = (\hat{\varphi} - \hat{\varphi}_{(i)})^T K(\varphi) (\hat{\varphi} - \hat{\varphi}_{(i)});$$

$\alpha = \frac{p}{p+3}$, siendo p el número de parámetros de β y el valor 3 es el número de parámetros de φ ;

$K(\beta)$ y $K(\varphi)$ son definidos en las ecuaciones (54) y (55) respectivamente;

$\hat{\beta}_{(i)}$ es el estimado de ML de β excluyendo la i-ésima observación;

$\hat{\varphi}_{(i)}$ es el estimado de ML de φ excluyendo la i-ésima observación.

4.2 Influencia local

Un problema que puede ocurrir en el estudio de la influencia global, con la eliminación individual de observaciones es la no detección de puntos conjuntamente influyentes en las estimación de los parámetros. Además, cuanto mayor cantidad de elementos de la muestra, mayor es el costo computacional exigido.

El método de influencia local propuesto por COOK (1986) consiste en evaluar, por medio de una medida apropiada de influencia, la robustez de las estimaciones dadas por el modelo, mediante pequeñas perturbaciones aplicadas al modelo o a los datos, o sea, verificar la existencia de puntos que sobre pequeñas modificaciones en el modelo, causen distorsiones en los resultados sobre los estimadores de ML, sin la necesidad de la eliminación del conjunto de datos. De esta forma, el método de influencia local verifica la influencia conjunta de observaciones en la estimación de los parámetros.

Sea el modelo espacial lineal definido en la ecuación (48), sobre la suposición de que el vector de errores aleatorios ε tiene distribución de probabilidad normal n-variada, con vector de media cero y matriz de covarianza Σ , esto es, $\varepsilon \sim N_n(\mathbf{0}, \Sigma)$, es posible ajustar un modelo de regresión espacial $\hat{\mathbf{Z}} = \mathbf{X}\hat{\boldsymbol{\beta}}$. En cuanto al análisis de los residuos se realiza para investigar problemas con el modelo ajustado, un análisis de diagnóstico de influencia local es realizado asumiendo el modelo como correcto e investigando la robustez de los resultados bajo pequeñas perturbaciones en los datos. Una observación es denominada influyente si produce alteraciones relevantes en los resultados del análisis cuando es excluida o sometida a una pequeña perturbación.

Considerando una perturbación aditiva en la variable respuesta $\mathbf{Z}_\omega = \mathbf{Z} + \boldsymbol{\omega}$, donde $\boldsymbol{\omega}$ es un vector perteneciente a un espacio de perturbación Ω , el logaritmo de la función de verosimilitud perturbada $l(\boldsymbol{\theta} | \boldsymbol{\omega})$, para el modelo normal es presentado como:

$$l(\boldsymbol{\theta}_\omega) = l(\boldsymbol{\theta} | \boldsymbol{\omega}) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log|\Sigma| - \frac{1}{2} (\mathbf{Z}_\omega - \mathbf{X}\boldsymbol{\beta})^T \Sigma^{-1} (\mathbf{Z}_\omega - \mathbf{X}\boldsymbol{\beta}). \quad (57)$$

La influencia de la perturbación $\boldsymbol{\omega}$ en el estimador de máxima verosimilitud del vector de parámetros $\boldsymbol{\theta}$ puede ser evaluada por la diferencia entre las verosimilitudes con y sin la perturbación, definida por la como:

$$LD(\boldsymbol{\omega}) = 2 \left(l(\hat{\boldsymbol{\theta}}) - l(\hat{\boldsymbol{\theta}}_\omega) \right), \quad (58)$$

donde, $\hat{\boldsymbol{\theta}}$ es el estimador de máxima verosimilitud de $\boldsymbol{\theta}$ del modelo postulado y $\hat{\boldsymbol{\theta}}_\omega$ es el estimador de máxima verosimilitud de $\boldsymbol{\theta}$ del modelo perturbado por $\boldsymbol{\omega}$.

COOK (1986) propuso estudiar el comportamiento local de $LD(\omega)$ alrededor de $\omega_0 \in \Omega$, tal que $l(\theta) = l(\theta|\omega_0)$. Para esto, se utiliza la curvatura normal C_l de $LD(\omega)$ en ω_0 la dirección de algún vector unitario l , definiendo en la ecuación (59) como:

$$C_l = 2 \| l^T \Delta^T L^{-1} \Delta l \|, \quad (59)$$

donde, $\| l \| = 1$; L : es la matriz hessiana, evaluada en $\theta = \hat{\theta}$; Δ : es una matriz, $(p+3) \times n$, dada por $\Delta = (\Delta_\beta^T, \Delta_\varphi^T)^T$, evaluada en $\theta = \hat{\theta}$ y en $\omega = \omega_0$, en que para la perturbación aditiva en la variable respuesta $Z_\omega = Z + \omega$, se tiene (URIBE-OPAZO et al., 2012):

$$\Delta_\beta = \frac{\partial^2 l(\theta|\omega)}{\partial \beta \partial \omega^T} = X^T \Sigma^{-1}; \quad \Delta_\varphi = \frac{\partial^2 l(\theta|\omega)}{\partial \varphi \partial \omega^T}, \text{ que tiene como elementos,}$$

$$\Delta_{\varphi j} = \frac{\partial^2 l(\theta|\omega)}{\partial \varphi_j \partial \omega^T} = \varepsilon^T \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_j} \Sigma^{-1}, j=1,2,3. \quad \text{con } \varepsilon = \varepsilon(\omega_0) = (Z - X\beta).$$

ZHU et al. (2007) propusieron una perturbación en la variable respuesta utilizando la matriz de covarianza, $Z_\omega = Z + \Sigma^{-1/2} \omega$, en que la matriz Δ dada por $\Delta = (\Delta_\beta^T, \Delta_\varphi^T)^T$ evaluada en $\theta = \hat{\theta}$ y en $\omega = \omega_0$. Las matrices Δ_β y Δ_φ son:

$$\Delta_\beta = X^T \Sigma^{-1} \Sigma^{-1/2}; \quad \Delta_\varphi = \frac{\partial^2 l(\theta|\omega)}{\partial \varphi \partial \omega^T}, \text{ que tiene como elementos}$$

$$\Delta_{\varphi j} = \varepsilon^T \Sigma^{-1} \left[\frac{\partial \Sigma}{\partial \varphi_j} \Sigma^{-1} + \Sigma^{-1/2} \frac{\partial \Sigma^{1/2}}{\partial \varphi_j} \right] \Sigma^{-1/2}, j=1,2,3, \text{ con } \varepsilon = \varepsilon(\omega_0) = (Z - X\beta).$$

DE BASTIANI et al. (2015) propusieron una perturbación generalizada $Z_\omega = Z + A\omega$, en base al trabajo de ZHU et al. (2007) en la variable respuesta, utilizando una matriz A , $n \times n$, que no depende de β y ω , tal que la matriz de información de Fisher para Z_ω con respecto al vector perturbado ω es de la forma $G(\omega_0) = cA^T \Sigma^{-1} A$, con c una constante positiva. En general $A^T \Sigma^{-1} A \neq I_n$, pero, si $A = \Sigma^{1/2}$, entonces $G(\omega_0) = cI_n$ siendo, la perturbación apropiada $Z_\omega = Z + \Sigma^{1/2} \omega$ para la variable respuesta. En este estudio las matrices Δ_β y Δ_φ evaluada en $\theta = \hat{\theta}$ y en $\omega = \omega_0$, son de la forma:

$$\Delta_{\beta} = -X^T \Sigma^{-1} A; \quad \Delta_{\varphi} = \frac{\partial^2 l(\theta | \omega)}{\partial \varphi \partial \omega^T}, \text{ tiene como elementos,}$$

$$\Delta_{\varphi j} = \varepsilon^T \left[\Sigma^{-1} \frac{\partial A}{\partial \varphi_j} - D_j A \right], \text{ para } j=1,2,3, \text{ con } \varepsilon = \varepsilon(\omega_0) = (Z - X\beta) \text{ y } D_j = \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_j} \Sigma^{-1}.$$

Las derivadas de primer y segundo orden de la matriz de covarianza Σ son dadas en las ecuaciones (60) a (66) (URIBE-OPAZO et al., 2012):

$$\frac{\partial \Sigma}{\partial \varphi_1} = I; \quad \frac{\partial^2 \Sigma}{\partial \varphi_1^2} = \frac{\partial^2 \Sigma}{\partial \varphi_1 \partial \varphi_2} = \frac{\partial^2 \Sigma}{\partial \varphi_1 \partial \varphi_3} = 0, \quad (60)$$

$$\frac{\partial \Sigma}{\partial \varphi_2} = R; \quad \frac{\partial^2 \Sigma}{\partial \varphi_2^2} = 0; \quad \frac{\partial^2 \Sigma}{\partial \varphi_2 \partial \varphi_3} = \frac{\partial R}{\partial \varphi_3} = \left[\left(\frac{\partial r_{ij}}{\partial \varphi_3} \right) \right], \quad (61)$$

$$\frac{\partial \Sigma}{\partial \varphi_3} = \varphi_2 \frac{\partial R}{\partial \varphi_3}; \quad \frac{\partial^2 \Sigma}{\partial \varphi_3^2} = \varphi_2 \frac{\partial^2 R}{\partial \varphi_3^2}. \quad (62)$$

En particular, para la función de covarianza exponencial se tiene:

$$\frac{\partial r_{ij}}{\partial \varphi_3} = \frac{r_{ij} h_{ij}}{\varphi_3^2}; \quad \frac{\partial^2 r_{ij}}{\partial \varphi_3^2} = \left(\frac{r_{ij} h_{ij}}{\varphi_3^3} \right) \left[\left(\frac{h_{ij}}{\varphi_3} \right) - 2 \right] \quad (63)$$

Para la función de covarianza gaussiana se tiene:

$$\frac{\partial r_{ij}}{\partial \varphi_3} = 2 \frac{r_{ij} h_{ij}^2}{\varphi_3^3}; \quad \frac{\partial^2 r_{ij}}{\partial \varphi_3^2} = \left(\frac{2 r_{ij} h_{ij}^2}{\varphi_3^4} \right) \left[2 \left(\frac{h_{ij}}{\varphi_3} \right)^2 - 3 \right] \quad (64)$$

Para la función de covarianza de la familia Matérn se tiene:

$$\frac{\partial r_{ij}}{\partial \varphi_3} = - \left(\frac{1}{\varphi_3} \right) \left[k r_{ij} + \frac{1}{2^{k-1} \Gamma(k)} \left(\frac{h_{ij}}{\varphi_3} \right)^{k+1} K'_k \left(\frac{h_{ij}}{\varphi_3} \right) \right], \quad (65)$$

$$\text{donde: } K'_k(u) = \frac{\partial K_k(u)}{\partial u} = \frac{1}{2} [K_{k-1}(u) + K_{k+1}(u)].$$

$$\frac{\partial^2 r_{ij}}{\partial \varphi_3^2} = \left(\frac{k(k+1) r_{ij}}{\varphi_3^2} \right) + \left(\frac{1}{\varphi_3^2 2^{k-1} \Gamma(k)} \right) \left(\frac{h_{ij}}{\varphi_3} \right)^{k+1} \left[2(k+1) K'_k \left(\frac{h_{ij}}{\varphi_3} \right) + \left(\frac{h_{ij}}{\varphi_3} \right) K''_k \left(\frac{h_{ij}}{\varphi_3} \right) \right], \quad (66)$$

$$\text{donde, } K''_k(u) = \frac{\partial^2 K_k(u)}{\partial u^2} = \frac{1}{4} [K_{k-2}(u) + 2K_k(u) + K_{k+2}(u)].$$

Considerando la matriz $\mathbf{B} = \mathbf{\Delta}^T \mathbf{L}^{-1} \mathbf{\Delta}$ y el elemento $C_i = 2|b_{ii}|$, para $i=1, \dots, n$, donde b_{ii} es el elemento de la diagonal principal de la matriz \mathbf{B} . Se puede utilizar el gráfico C_i versus i (orden de los datos) como una técnica de diagnóstico para evaluar la existencia de observaciones influyentes. VERBEKE & MOLENBERGHS (2001) proponen considerar como potencialmente influyentes la i -ésima observación si, $C_i > 2\bar{C}$, donde $\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i$.

Sea L_{max} el primer autovector, normalizado y asociado a los mayor autovalor, en módulo, de la matriz $\mathbf{B} = \mathbf{\Delta}^T \mathbf{L}^{-1} \mathbf{\Delta}$. El gráfico de los elementos $|L_{max}|$ versus i (orden de los datos) puede destacar los elementos que, sobre pequeñas perturbaciones, ejercen mayor influencia en $LD(\omega)$, en una vecindad de ω_0 (COOK, 1986).

4.3 Influencia local conformal

POON & POON (1999) observan que la curvatura normal propuesta por COOK (1986) puede asumir cualquier valor y no es invariante sobre una transformación uniforme de escala. A fin de resolver este problema, los autores definen una nueva medida de curvatura que tiene una interpretación más simple con buenas propiedades, que llamaron curvatura normal conformal. Esa curvatura en la dirección de un vector \mathbf{l} evaluada en ω_0 es dada por:

$$B_l = \frac{C_l}{tr(2\mathbf{B})}$$

Una interesante propiedad de la curvatura normal conformal es que para cualquier dirección B_l satisface la condición que $0 \leq B_l \leq 1$. Esto permite comparar la curvatura normal en diferentes modelos.

Para la comparación de las curvaturas de diferentes modelos, de acuerdo con ZHU & LEE (2001), el i -ésimo elemento es potencialmente influyente si, $B_i > \bar{B} + 2DP(B)$, siendo $\bar{B} = \frac{1}{n} \sum_{i=1}^n B_i$ y $DP(B)$ es la desviación estándar de B_1, \dots, B_n .

CAPÍTULO 5

APLICACIONES

En este trabajo se realizaron estudios de dependencia espacial para variables regionalizadas en dos áreas experimentales con tamaño y características de suelo diferentes en la cual el acompañamiento de la producción es realizada mediante la agricultura de precisión.

5.1 Caracterización del área I en estudio

Los datos fueron obtenidos por los investigadores del Laboratorio de Estadística Espacial - LEE/UNIOESTE, en el año agrícola 2004/2005, en una área de Latos solo Rojo Distroférico, de textura arcillosa (EMBRAPA, 2009) con 57 hectáreas (ha), ubicado en el municipio de Cascavel – PR, Brasil, con coordenadas geográficas de 24,95° sur de latitud y 53,57° oeste de longitud, con altitud media de 650 m. El clima de la región es templado mesotérmico y muy húmedo, tipo climático Cfa (Köppen) y la temperatura anual media de 21 °C. El levantamiento topográfico y la verificación del posicionamiento de los locales de muestreo fueron realizados por medio del receptores GPS, GEOEXPLORE 3, en un sistema espacial de coordenadas *Universal Transverse Mercatur* (UTM). La cuadrícula (*"grid"*er) muestral es regular espacialmente georreferenciada, fue construida con una densidad muestral de 75 x 75 m de distancia entre los puntos de la muestra, se obteniendo 93 elementos como es representado gráficamente en la Figura 5.

La variable regionalizada en estudio fue la resistencia del suelo a penetración (RSP) en Mega Pascal [Mpa] en la capa de 0.0 a 0.10 m de profundidad, medida con un penetrógrafo

de marca SOILCONTROL SC-60, con asta de 600 mm y 9.53 mm de diámetro, equipado en la punta con un cono de 129.3 mm² de área de base, 12.83 mm de diámetro y 30 grados de ángulo de vértice, siendo realizadas cuatro repeticiones por local, realizadas de forma aleatoria en cada parcela experimental, definido un valor medio de resistencia del suelo a la penetración. La capa 0.0 - 0.1 m de profundidad, tenía una humedad media de 12.74%.

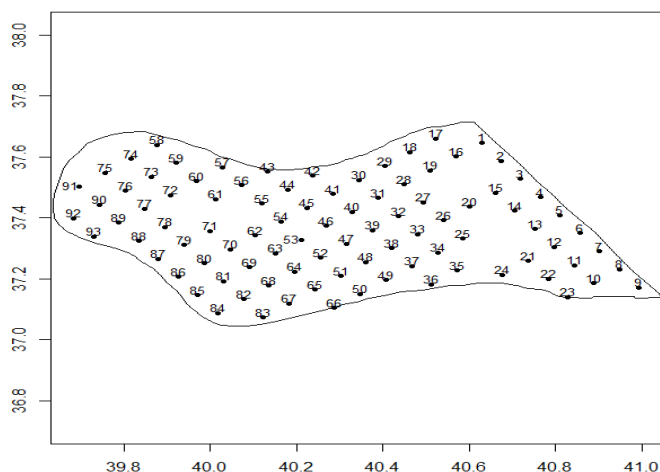


Figura 5. Esquema muestral del área I en estudio con una cuadrícula regular de 75 x 75 m.

5.2 Caracterización del área II en estudio

Los datos de Productividad de trigo de la aplicación II fueron colectados por investigadores en un experimento realizado en el año agrícola de 2003, junto a la Cooperativa Central de Pesquisa Agrícola (COODETEC), situada en la BR 467, km 98, en el municipio, en la región oeste del estado del Paraná-Brasil, longitud 53° 26' W y latitud 25° 56' S. El suelo fue clasificado como Lato solo Rojo Distroférico, de textura arcillosa (EMBRAPA, 2009) y fue descompactado el suelo antes de la plantación. En la Figura 6 se representa el área experimental.

El área total del experimento corresponde a 22.62 hectáreas (ha), en la cual fueron plantadas dos variedades de trigo: COODETEC 103 (CD 103) y COODETEC 101 (CD 101). El área total correspondiente al cultivo de trigo de la variedad CD 103 fue de 18.17 ha, siendo

área 2 con 11.06 hectáreas y el área 3 con 7.11 hectáreas (Figura 6). La variedad CD 101 de trigo fue cultivada en 4.45 hectáreas (área 1). La selección de la variedad de trigo y de las áreas utilizadas fueron realizadas por determinación de los técnicos del centro de pesquisa al cual las áreas pertenecen, siguiendo intereses propios. La cuadrícula muestral regular considerada en el estudio fue de 50 x 50 m, totalizando una muestra de 83 elementos.

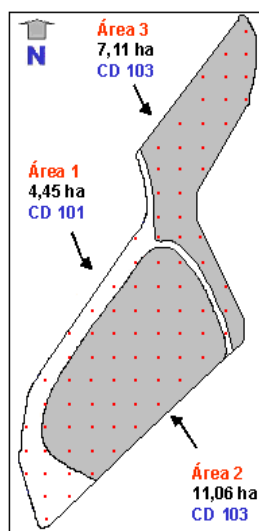


Figura 6. Esquema muestral del área II en estudio, con una cuadrícula regular de 50 x 50 m.

5.3 Análisis de los datos

Inicialmente, fue realizado el análisis exploratorio de los datos de Productividad de trigo para evaluar el comportamiento general de cada variable regionalizada e identificar la posible existencia de puntos atípicos con su ubicación en el área experimental utilizando el gráfico *boxplot*. También fue verificada la existencia de tendencia direccional por medio del gráfico *postplot*. Seguidamente, se verificó la presencia o no de anisotropía por medio de las semivarianzas obtenidas en las diferentes direcciones en estudio que son: 0°, 45°, 90° e 135°.

El siguiente paso fue el análisis espacial de los datos utilizando los métodos geoestadísticos. Para la identificación de la estructura de dependencia espacial fue utilizado el ajuste de los modelos teóricos exponencial, gaussiano y de la familia Matérn, siendo los parámetros de los modelos estimados por máxima verosimilitud. En esta etapa, fueron

utilizados los criterios de validación cruzada, *AIC* e *SIC* para escoger el mejor modelo ajustado y construir los mapas temáticos por la técnica de kriging universal para las variables regionalizadas en estudio.

Por las técnicas de diagnósticos de influencia global y local en el modelo espacial lineal gaussiano, se verificó la existencia de observaciones que pueden ejercer influencia en el desplazamiento de la verosimilitud cuando se tiene diferentes tipos de perturbación en la variable respuesta. Con la identificación de los puntos influyentes fueron retirados los puntos del conjunto de datos total y fueron realizadas nuevas análisis de dependencia espacial identificando en que aspecto los puntos considerados influyentes afectan la análisis espacial y consecuentemente afectan la construcción de los mapas temáticos. Los mapas temáticos contruidos con todos los puntos y sin los puntos considerados influyentes fueron comparados, utilizando medidas de exatitud, obtenidas a partir de la matriz de errores, construida con los datos interpolados por kriging con todos los puntos y sin los puntos considerados influyentes.

5.4 Software utilizado

Para el análisis de los datos fue utilizado el *software* R (R DEVELOPMENT CORE TEAM, 2013), versión 3.0.1 y módulo geoR (RIBEIRO JR & DIGGLE, 2013).

CAPÍTULO 6

RESULTADOS Y DISCUSIONES

6.1 Análisis del área I: Estudio de la resistencia del suelo a la penetración - RSP

6.1.1 Análisis exploratorios de datos

En la Tabla 2, se presentan los resultados de las estadísticas descriptivas para la variable resistencia del suelo a la penetración - RSP con 93 puntos de la muestra analizados (los datos son presentados en el Anexo A). El valor medio de la RSP en la capa de 0.0 - 0.10 m de profundidad es de 2.729 Mpa (Mega Pascal) y es considerado un valor alto, según las recomendaciones de CANARACHE (1994), que indican que valores de la RSP en el intervalo [2.6; 5.0] Mpa representan suelos con limitaciones en el crecimiento de las raíces. Se observa un bajo valor de la desviación estándar de los datos de la RSP en relación a su valor medio (DP = 0.669 Mpa) y el coeficiente de variación (CV = 24.49%), que indica homogeneidad de los datos (GOMES, 2009). Según los resultados de los coeficientes de asimetría y curtosis muestral, los datos de la RSP en la capa 0.0 - 0.10 m de profundidad presentan una distribución con características similares a los de una distribución normal, de acuerdo con los intervalos $[-0.488; 0.488]$ y $[-0.720; 1.150]$ de 95% de confianzas presentados en JONES (1969) para estas dos estadísticas, respectivamente.

Tabla 2. Estadísticas descriptivas de la RSP (Mpa) en la capa 0.0 - 0.10 m de profundidad

| Variable | n | Mín | Media | Máx. | Q1 | Mediana | Q3 | DP | CV(%) | Ass | Curt |
|----------|----|-------|-------|-------|-------|---------|-------|-------|-------|--------|-------|
| RSP | 93 | 0.671 | 2.729 | 4.269 | 2.374 | 2.692 | 3.125 | 0.669 | 24.49 | -0.169 | 0.516 |

n: número de datos; Mín: valor mínimo; Máx.: valor máximo; Q1: primer cuartil; Q3: tercer cuartil; Me: Mediana; DP: desviación estándar; CV: coeficiente de variación; Ass: Coeficiente de asimetría; Curt: Coeficiente de curtosis.

En las Figuras 7 y 8, se presentan los *boxplot* y *postplot*, respectivamente, para la RSP en la capa 0.0 a 0.10 m de profundidad. Se puede observar que en gráfico *boxplot* de la Figura 7, la distribución de los datos es simétrica y existen cuatro observaciones discrepantes (*outliers*)

que son: 01, 55, 89 y 90. En el gráfico *postplot* de la Figura 8, identifican las ubicaciones de estos cuatro puntos discrepantes en el área en estudio y se observa que existen altos valores de la RSP en la región extrema (Este).

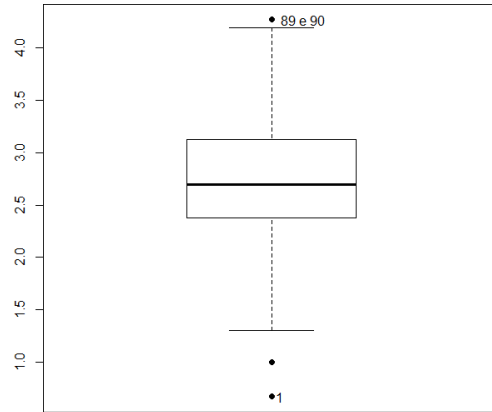


Figura 7. Gráficos *boxplot* de la RSP en la capa 0.0 - 0.10 m de profundidad.

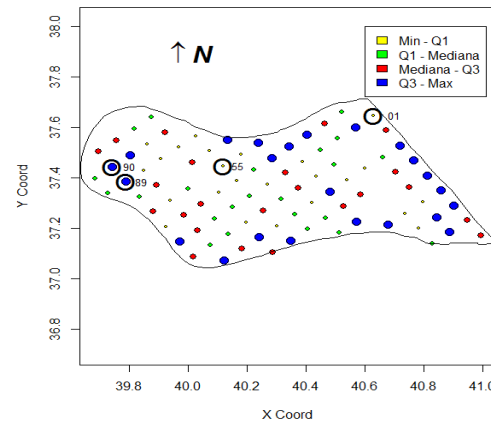


Figura 8. Gráfico *postplot* de la RSP en la capa 0.0 - 0.10 m de profundidad.

En la Figura 9a se presentan los gráficos de las funciones semivarianza muestral en las direcciones 0° , 45° , 90° y 135° . Se observan la similitud de las funciones semivarianzas en estas cuatro direcciones hasta una distancia de 700 m, que corresponde al 50% de la distancia máxima (1337.63 m) entre puntos (*cutoff*), distancia recomendada por CLARK (1979). De esta forma, se puede afirmar que el comportamiento de la estructura de dependencia espacial es semejante en las cuatro direcciones en estudio, lo que indican que los datos son isotrópicos, esto es, que no existen direcciones privilegiadas. De esta manera, no es necesaria la incorporación de la anisotropía en el modelo, y por lo tanto, es utilizado un solo semivariograma llamado semivariograma omnidireccional muestral para el estudio de la variabilidad espacial (GUEDES et al., 2008), como es presentado en la Figura 9b.

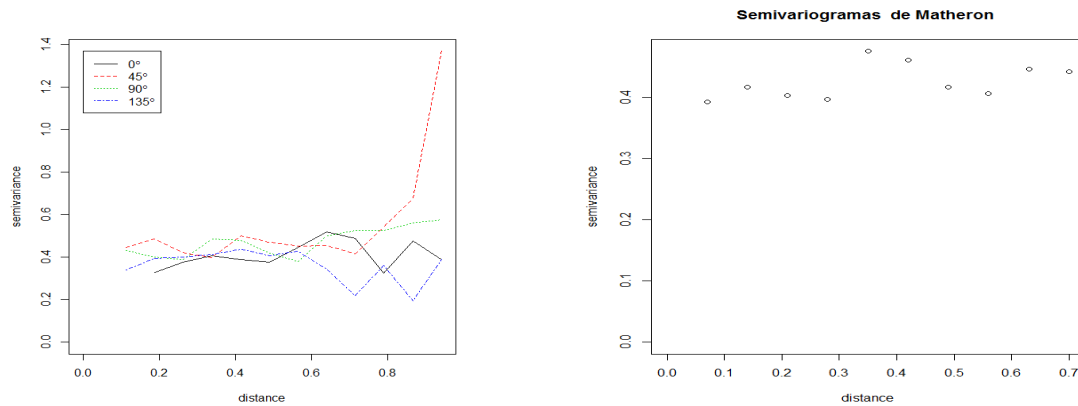


Figura 9. Gráfico de los semivariogramas direccionales de la RSP (a); Gráfico del semivariograma omnidimensional muestral de la RSP (b).

En la Tabla 3 se presenta la información sobre el semivariograma omnidimensional muestral hasta una distancia de 700 m, que corresponde al 50% de la distancia máxima entre puntos (*cutoff*). En esta tabla se presentan las distancias entre los “lags”, valor de la función semivarianza estimada de Matheron y el número de pares utilizados en el cálculo de la función de semivarianza.

Tabla 3. Información sobre el semivariograma omnidimensional muestral de Matheron para la RSP (Mpa) en la capa 0.0-0.10 m de profundidad

| | | | | | | | | |
|-------------------------|--------|--------|--------|--------|--------|--------|--------|--------|
| Distancia (h) km | 0.0836 | 0.1673 | 0.2509 | 0.3345 | 0.4181 | 0.5018 | 0.5854 | 0.6690 |
| Semivarianza | 0.3886 | 0.4289 | 0.3942 | 0.4560 | 0.4552 | 0.4142 | 0.4075 | 0.4745 |
| Número de pares N(h) | 281 | 347 | 572 | 658 | 541 | 466 | 349 | 337 |

6.1.2 Análise de la estructura de dependencia espacial y criterios de validación

En la Tabla 4 son presentados los resultados de los análisis de la estructura de dependencia espacial para la variable RSP en la capa 0.0-0.10 m de profundidad, en que fueron ajustados varios modelos teóricos, siendo los parámetros estimados por el método de máxima verosimilitud (ML). Se observa que en todos los modelos ajustados, existen diferencias entre las estimaciones de $\hat{\phi}_2$ y $\hat{\phi}_3$, siendo que en todos los modelos estudiados las mesetas ($\hat{\phi}_1 + \hat{\phi}_2$) son similares. Los valores estimados para el alcance (\hat{a}) que define el radio de dependencia espacial de cada modelo varían entre 108.4 a 151.0 m. Se puede decir que

según el coeficiente de efecto pepita relativo $E = (\hat{\phi}_1/(\hat{\phi}_1 + \hat{\phi}_2))100\%$, existe fuerte dependencia espacial ($E \leq 25\%$) y moderada dependencia espacial ($25\% < E < 75\%$) en los modelos ajustados (CAMBARDELLA et al., 1994). Finalmente, se observa que el modelo gaussiano es el que presenta mayor valor del logaritmo de la función de verosimilitud (MLL).

Tabla 4. Análisis de la estructura de dependencia espacial para la variable RSP, utilizando el método de estimación ML

| Método | Modelo | $\hat{\beta}$ | $\hat{\phi}_1$ | $\hat{\phi}_2$ | $\hat{\phi}_3$ | \hat{a} (m) | E (%) | MLL |
|--------|-----------------|--------------------------------|--------------------------------|--------------------------------|----------------------------------|---------------|--------------|----------------|
| ML | Exponencial | 2.741 (0.088) | 0.000 (0.306) | 0.443 (0.328) | 0.0362 (0.00004) | 108.4 | 0.00 | -92.95 |
| | Gausiano | 2.739 (0.089) | 0.317 (0.083) | 0.127 (0.088) | 0.087 (0.00039) | 151.0 | 71.40 | - 92.83 |
| | Matérn $k=0.7$ | 2.741 (0.088) | 0.955 (0.236) | 0.348 (0.257) | 0.0341 (0.00004) | 117.0 | 73.28 | - 92.94 |
| | Matérn $k=1.0$ | 2.741 (0.089) | 0.170 (0.183) | 0.273 (0.201) | 0.030 (0.00004) | 126.0 | 38.35 | -92.94 |
| | Matérn $k=1.5$ | 2.741 (0.089) | 0.220 (0.149) | 0.230 (0.165) | 0.030 (0.00003) | 132.0 | 48.78 | -92.93 |

$\hat{\beta}$: media estimada de los datos; $\hat{\phi}_1$: efecto pepita estimado; $\hat{\phi}_2$: contribución estimada; $\hat{\phi}_3$: función del alcance estimada; \hat{a} : alcance estimado (m); E: efecto pepita relativo (%), $E = \hat{\phi}_1/(\hat{\phi}_1 + \hat{\phi}_2) \times 100$; MLL: máximo valor del logaritmo de la función de verosimilitud, entre paréntesis son presentados las desviaciones estándares.

En la Tabla 5 son presentados los resultados de los criterios de validación cruzada, AIC y SIC para seleccionar el mejor modelo ajustado. Según los criterios analizados y conjuntamente con el valor del logaritmo de la función de verosimilitud MLL (Tabla 4), el mejor modelo ajustado es el gaussiano, con parámetros estimados $\hat{\beta} = 2.739$, $\hat{\phi}_1 = 0.317$, $\hat{\phi}_2 = 0.127$, $\hat{\phi}_3 = 0.087$ y con un radio de dependencia espacial de 151.0 m.

Tabla 5. Criterios de validación cruzada, Akaike y Schwarz en la selección del mejor ajuste

| Método de estimación | Modelo | EM | \overline{ER} | S_{EM} | S_{MR} | EA | AIC | SIC |
|----------------------|-----------------|-----------------|-----------------|----------------|----------------|---------------|--------------|--------------|
| ML | Exponencial | 0.000523 | 0.00040 | 0.65826 | 1.00912 | 46.229 | 193.9 | 204.0 |
| | Gausiano | 0.000560 | 0.00043 | 0.65608 | 1.00918 | 45.932 | 193.7 | 203.8 |
| | Matérn $k=0.7$ | 0.000525 | 0.000403 | 0.658173 | 1.009158 | 46.217 | 193.9 | 204.0 |
| | Matérn $k=1.0$ | 0.000527 | 0.000404 | 0.658038 | 1.009151 | 46.197 | 193.9 | 204.0 |
| | Matérn $k=1.5$ | 0.000534 | 0.000410 | 0.657893 | 1.009349 | 46.184 | 193.9 | 204.0 |

EM: error medio; \overline{ER} : error medio reducido; S_{EM} : desviación estándar del error medio; S_{ER} : desviación estándar del error reducido; EA: error absoluto; AIC: criterio de información de Akaike; SIC: criterio de información de Schwarz.

6.1.3 Construcción de mapas temáticos

Por medio de la interpolación por kriging ordinario, considerando la estructura de dependencia espacial del modelo gaussiano, que mejor se ajustó a la función semivarianza, fue obtenido el mapa temático para la RSP en la capa 0 - 0.10 m de profundidad, presentada en la Figura 10. Se verifican sectores con grandes diferencias de la RSP en el área en estudio, variando gran parte del área de 2.54 a 2.79 Mpa y en los sectores extremos inferiores variando de 3.03 a 3.27 Mpa.

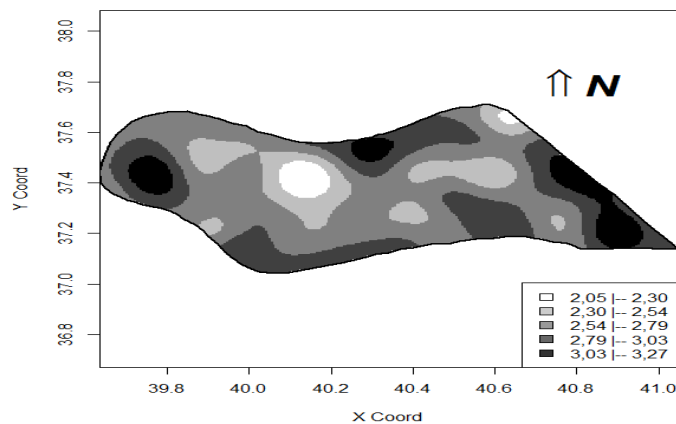


Figura 10. Mapa temático de la RSP [Mpa] en la capa 0.0 - 0.10 m de profundidad.

6.1.4 Diagnósticos de Influencia local

En el estudio de diagnóstico de influencia local para los datos de RSP fueron utilizados los gráficos C_i versus orden y $|L_{max}|$ versus orden, de acuerdo con el modelo ajustado, con la finalidad de evaluar si algunas de las observaciones de la muestra estarían ejerciendo algún tipo de influencia en el desplazamiento de la verosimilitud. El estudio de influencia local fue realizado para la variable respuesta, utilizando los esquemas de perturbación aditiva, perturbación de ZHU et al. (2007) y perturbación de Zhu generalizada propuesto en DE BASTIANI et al. (2014) y que están presentadas en la Sección 4.2.

En las Figuras 11 a 16 son presentados los resultados gráficos de los diagnósticos de influencia local para los tres esquemas de perturbación de la variable respuesta. Los gráficos

presentados son de los coeficientes C_i versus el orden de las observaciones recolectadas según la localización y los gráficos $|L_{max}|$ versus a orden de las observaciones recolectadas. Por el análisis gráfico de las Figuras 11 a 14, se verifican en los estudios de los esquemas de perturbación aditiva y Zhu que existen tres observaciones 01, 85 y 86 consideradas como influyentes en el desplazamiento de la verosimilitud. Al utilizar el esquema de perturbación de Zhu generalizado se observan en los gráfico C_i versus orden (Figura 15) las observaciones 54, 61 y 87 son consideradas influyentes y para el gráfico $|L_{max}|$ versus orden (Figura 16) las observaciones 07, 61 y 92 son consideradas influyentes.

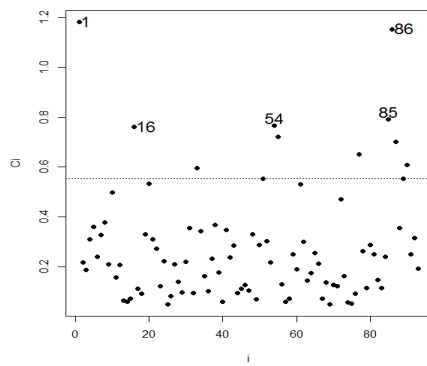


Figura 11. Gráfico C_i versus a orden, de la RSP utilizando la perturbación aditiva

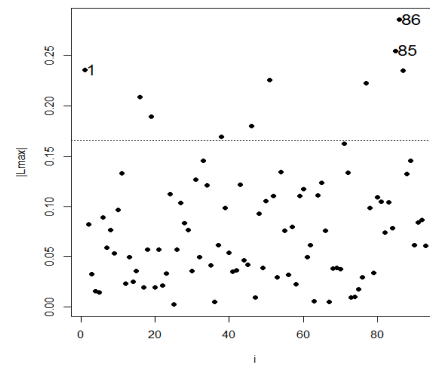


Figura 12. Gráfico $|L_{max}|$ versus orden, de la RSP utilizando la perturbación aditiva.

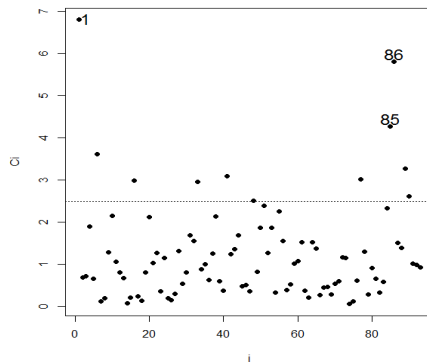


Figura 13. Gráfico C_i versus a orden, de la RSP, utilizando la perturbación de Zhu.

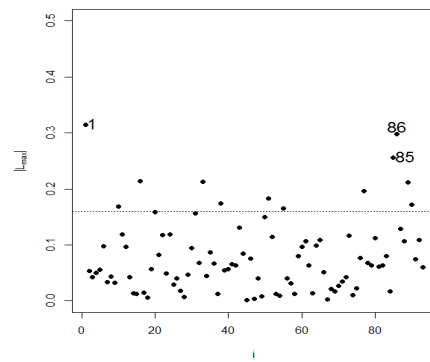


Figura 14. Gráfico $|L_{max}|$ versus orden, de la RSP, utilizando la perturbación de Zhu.

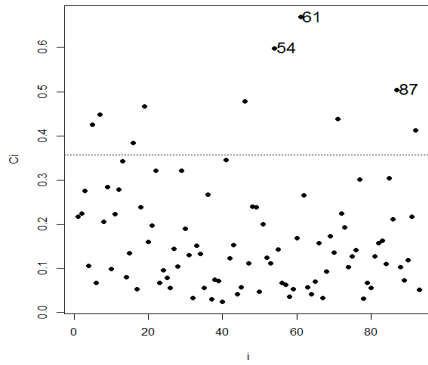


Figura 15. Gráfico C_i versus a orden, de la RSP utilizando la perturbación de Zhu generalizada.

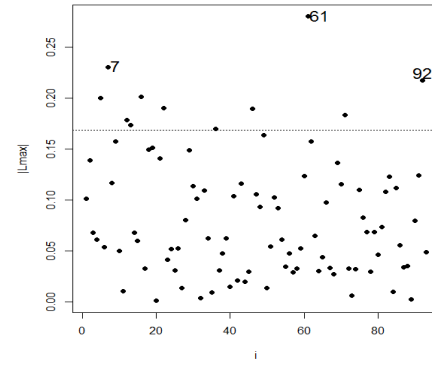


Figura 16. Gráfico $|L_{max}|$ versus orden, de la RSP utilizando la perturbación de Zhu generalizada.

Para presentar los efectos de las observaciones influyentes en el análisis de la estructura de variabilidad espacial y en la construcción de los mapas temáticos, fue retirada la observación 01 considerada influyente por los esquemas de perturbación aditiva y Zhu, realizando posteriormente un nuevo estudio de variabilidad espacial.

6.1.5 Análisis de la RSP sin considerar la observación influyente

En la Tabla 6 se presentan los resultados de las estadísticas descriptivas para la variable RSP en la capa 0 – 0.10 m de profundidad sin la observación 01 considerada influyente, según los esquemas de perturbaciones aditiva y Zhu. Se verifica que el valor medio de la RSP aumentó a 2.752 Mpa, siendo aún considerado alto para la RSP, de acuerdo con las recomendaciones de CANARACHE (1994) indicando la existencia de algunas limitaciones en el crecimiento de las raíces. Se verifica la disminución de la desviación estándar de los datos de la RSP en relación a la media (0.636 Mpa) y la disminución del coeficiente de variación ($CV = 23.12\%$). Según los coeficientes de asimetría y curtosis, la distribución de la RSP en la capa 0.0-0.10 m de profundidad continúa presentando características de la distribución normal de probabilidades de acuerdo con los intervalos de 95% de confianza presentados en JONES (1969).

Tabla 6. Estadísticas descriptivas de la RSP (Mpa) en la capa 0.0-0.10 m de profundidad sin el punto 01 considerado influyente

| Variable | n | Mín. | Media | Máx. | Q1 | Mediana | Q3 | DP | CV(%) | Ass | Curt |
|----------|----|-------|-------|-------|-------|---------|-------|-------|-------|-------|-------|
| RSP | 92 | 0.999 | 2.752 | 4.269 | 2.384 | 2.694 | 3.126 | 0.636 | 23.12 | 0.066 | 0.124 |

n: número de datos; Mín.: valor mínimo; Máx.: valor máximo; Q1: primer cuartil; Q3: tercer cuartil; Me: Mediana; DP: desviación estándar; CV: coeficiente de variación; Ass: Coeficiente de asimetría; Curt: Coeficiente de curtosis.

El gráfico *postplot* de la Figura 17 identifica la ubicación donde la observación 01 considerada influyente, actuaba en la interpolación por kriging.

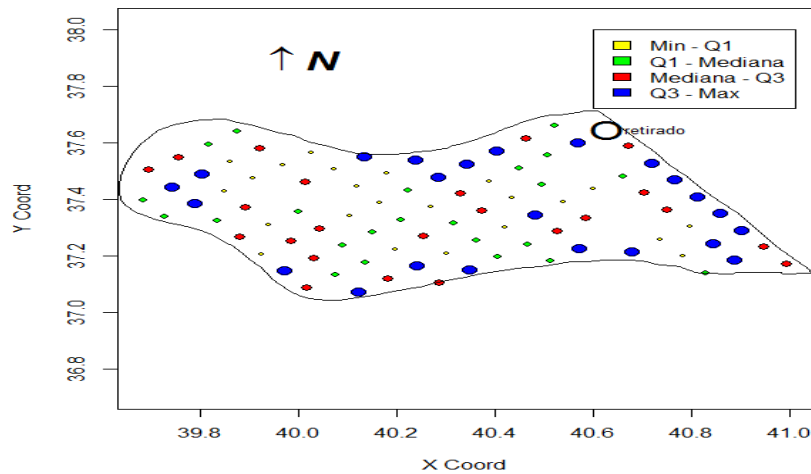


Figura 17. Gráfico *postplot* de la RSP en la capa 0.0 - 0.10 m de profundidad sin la observación 01 considerado influyente.

En la Tabla 7 son presentados los resultados de los análisis estadísticos espaciales para la variable RSP, en que fueron ajustados modelos teóricos, con parámetros estimados por máxima verosimilitud (ML) y sus respectivas desviaciones estándares. Se observa que sin la presencia de la observación 01 considerada influyente, existen cambios en las estimaciones $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\phi}_3$ y hubo disminución de las desviaciones estándares de las estimaciones.

En la Tabla 8 se presentan los criterios de validación cruzada, *AIC* y *SIC* para seleccionar el mejor modelo ajustado. Según los criterios utilizados, el mejor modelo ajustado fue el exponencial con parámetros estimados $\hat{\beta} = 2.775$, $\hat{\phi}_1 = 0.0$, $\hat{\phi}_2 = 0.401$, $\hat{\phi}_3 = 0.04$ y con radio de dependencia espacial de 121.0 m. Se observa que con la ausencia de la observación 01 considerada influyente el modelo de variabilidad espacial cambió de gaussiano para exponencial, variando el radio de dependencia espacial de 151.0 m para 121.0 m.

Tabla 7. Análisis de la estructura de dependencia espacial para la RSP sin el punto 01 considerado influyente, utilizando o método de estimación de parámetros ML

| Método de estimación | Modelo | $\hat{\beta}$ | $\hat{\phi}_1$ | $\hat{\phi}_2$ | $\hat{\phi}_3$ | \hat{a} (m) | E (%) | MLL |
|----------------------|--------------|-------------------------|------------------------|-------------------------|---------------------------|---------------|-------------|---------------|
| ML | Exponencial | 2.775 (0.089) | 0.00 (0.228) | 0.401 (0.253) | 0.040 (0.00004) | 121.0 | 0.00 | -86.68 |
| | Gausiano | 2.772 (0.089) | 0.269 (0.070) | 0.134 (0.079) | 0.090 (0.00038) | 155.0 | 71.40 | -86.52 |
| | Matérn k=0.7 | 2.775 (0.089) | 0.073 (0.182) | 0.328 (0.205) | 0.037 (0.00005) | 128.0 | 73.28 | -86.67 |
| | Matérn k=1.0 | 2.774 (0.089) | 0.138 (0.143) | 0.263 (0.164) | 0.026 (0.00004) | 136.0 | 38.35 | -86.66 |
| | Matérn k=1.5 | 2.774 (0.089) | 0.166 (0.115) | 0.214 (0.133) | 0.030 (0.00003) | 144.0 | 48.78 | -86.65 |

$\hat{\beta}$: media estimada dos datos; $\hat{\phi}_1$: efecto pepita estimado; $\hat{\phi}_2$: contribución estimada; $\hat{\phi}_3$: función del alcance estimado; \hat{a} : alcance estimado (m); E: efecto pepita relativo (%), $E = \hat{\phi}_1 / (\hat{\phi}_1 + \hat{\phi}_2) \times 100$; MLL: máximo valor del logaritmo de la función de verosimilitud, entre paréntesis son presentados los desviaciones estándares.

Tabla 8. Criterios de validación cruzada, Akaike- *AIC* y Schwarz-*SIC* para seleccionar el mejor modelo ajustado

| Modelo | <i>EM</i> | \overline{ER} | S_{EM} | S_{ER} | <i>EA</i> | <i>AIC</i> | <i>SIC</i> |
|--------------------|-----------------|-----------------|------------------|------------------|-----------------|--------------|--------------|
| Exponencial | 0.001243 | 0.001014 | 0.6159691 | 1.0071597 | 43.80034 | 181.4 | 191.4 |
| Gausiano | 0.001344 | 0.001100 | 0.6133028 | 1.0071792 | 43.49435 | 181.0 | 191.1 |
| Matérn k=0.7 | 0.001254 | 0.001023 | 0.6158871 | 1.0072939 | 43.79662 | 181.3 | 191.4 |
| Matérn k=1 | 0.001259 | 0.001028 | 0.6157112 | 1.0072957 | 43.7739 | 181.3 | 191.4 |
| Matérn k=1.5 | 0.001267 | 0.001034 | 0.6154706 | 1.0072775 | 43.74477 | 181.3 | 191.4 |

EM: error medio; \overline{ER} : error medio reducido; S_{EM} : desviación estándar de error medio; S_{ER} : desviación estándar del error medio reducido; *EA*: error absoluto; *AIC*: criterio de información de Akaike; *SIC*: criterio de información de Schwarz.

En la Figura 18 es presentado el mapa temático de la RSP en la capa de 0.0-0.10 m de profundidad, sin la observación 01 considerada influyente. El mapa fue construido por kriging ordinaria. Se verifican diferencias en la área en estudio con mayor RSP en los extremos debido al tráfico de máquinas agrícolas y menos RSP en la área central. En relación al mapa de la RSP construido con todos los puntos (Figura 10), se observan diferencias en los mapas temáticos.

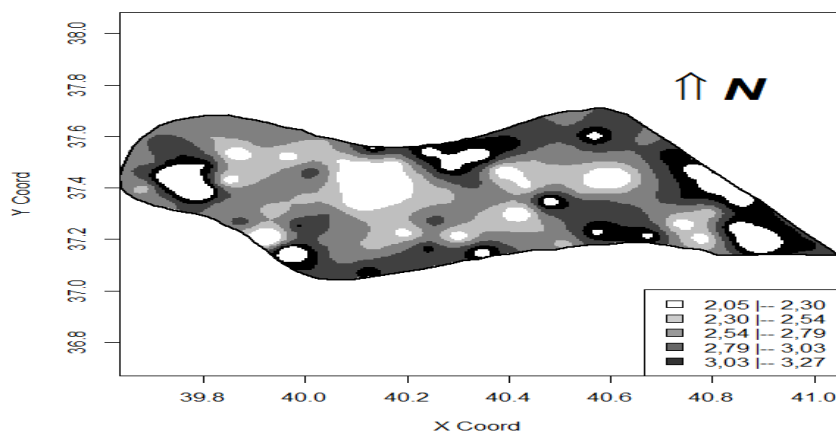


Figura 18. Mapa temático de la RSP en la capa 0.0 - 0.10 m de profundidad sin la observación 01 considerada influyente.

En las Tablas 9 y 10 son presentadas las distribuciones de frecuencia del número de *pixeles* y porcentajes de área por cada clase de la RSP, considerando el estudio con todos los puntos y considerando la eliminación de la observación 01 considerada influyente. Se observan diferencias en las distribuciones de los datos estimados por clase al ser retirada la observación 01 considerada influyente en relación a la distribución con todos los datos.

Tabla 9. Distribución del número de *pixeles* y porcentaje de área por clases para el mapa con todos los puntos

| Intervalo de Clases | Número <i>Pixels</i> | Área (ha) | Área % |
|---------------------|----------------------|-----------|--------|
| 2.05 -2.30 | 657 | 1.64 | 2.87 |
| 2.30 -2.54 | 3457 | 8.64 | 15.13 |
| 2.54 -2.79 | 10608 | 26.52 | 46.42 |
| 2.79 -3.03 | 6165 | 15.41 | 26.98 |
| 3.03 -3.27 | 1967 | 4.92 | 8.61 |
| Total | 22854 | 57.135 | 100.0 |

Tabla 10. Distribución del número de *pixeles* y porcentajes de área por clases para el mapa sin observación influyente

| Intervalo de Clases | Número <i>Pixels</i> | Área (ha) | Área % |
|---------------------|----------------------|-----------|--------|
| 2.05 -2.30 | 1845 | 4.61 | 8.07 |
| 2.30 -2.54 | 4319 | 10.80 | 18.90 |
| 2.54 -2.79 | 7152 | 17.88 | 31.29 |
| 2.79 -3.03 | 5208 | 13.02 | 22.79 |
| 3.03 -3.27 | 2330 | 10.83 | 18.95 |
| Total | 22854 | 57.135 | 100.0 |

En la Tabla 11 se presenta la matriz de errores para el estudio de comparación de mapas temáticos con todos los datos (referencia) y sin la observación 01 considerada influyente (modelo) de la RSP en la capa 0 - 0.10 m de profundidad.

En la matriz de errores de la Tabla 11, se cuantifica el número de “*pixels*” por clases del mapa de la RSP con todos los puntos (referencia) y del mapa de la RSP sin la observación 01

considerada influyente (modelo), donde en la diagonal principal se presenta el número de *pixeles* clasificados en común por ambos mapas. El índice de exactitud global (*EG*) calculado para la comparación de ambos mapas fue de $EG = 0.6459$, el cual es considerado según la clasificación de ANDERSON *et al.* (2001) de no existir semejanza entre ambos mapas. El índice Kappa calculado fue de $K = 0.5255$ indicando que existen poca similaridad entre los mapas temáticos interpolados por la kriging con y sin la observación 01 considerada influyente, la misma conclusión es obtenida con el índice Tau que fue $T = 0.5574$ (KRIPENDORFF, 2004). Finalmente, se puede concluir que la observación 01 ejerce una influencia desproporcional en la construcción del mapa temático, pudiendo cambiar el comportamiento de la variabilidad espacial de la RSP en la capa 0- 0.10 m de profundidad de la área en estudio.

Tabla 11. Matriz de errores para la comparación de mapas temáticos de la RSP en la capa 0 - 0.10 m de profundidad, por número de *pixeles*

| | Clases | Mapa temático con todos los puntos (referencia) | | | | | Total |
|------------------------------------------------|-------------|-------------------------------------------------|-------------|-------------|-------------|-------------|-------|
| | (MPa) | 2.05 --2.30 | 2.30 --2.54 | 2.54 --2.79 | 2.79 --3.03 | 3.03 --3.27 | |
| Mapa temático sin el punto influyente (modelo) | 2.05 --2.30 | 528 | 967 | 295 | 1 | 0 | 1845 |
| | 2.30 --2.54 | 0 | 2134 | 2162 | 23 | 0 | 4319 |
| | 2.54 --2.79 | 0 | 109 | 6569 | 474 | 0 | 7152 |
| | 2.79 --3.03 | 75 | 235 | 1387 | 3511 | 0 | 5208 |
| | 3.03 --3.27 | 0 | 12 | 195 | 2156 | 1969 | 4330 |
| Total | | 657 | 3457 | 10608 | 6165 | 1967 | 22854 |

6.2 Análisis del Área II: Estudio de la Productividad de trigo

6.2.1 Modelo espacial lineal

En el año agrícola 2003 fueron colectados datos de la productividad de trigo en tonelada por hectárea ($t\ ha^{-1}$), en una área de 22.62 hectáreas, en la cual fueron plantadas las variedades de trigo COODETEC 101 (CD 1013) y COOETEC 103 (CD 103). Las áreas correspondientes a variedad CD 101 fue de 4.45 ha y a la variedad CD 103 fue de 18.17 ha.

El principal interés fue estudiar la variabilidad espacial de la productividad de trigo en la área en estudio, considerando la variedad de trigo (V) como una variable explicativa Dummy

donde, $V(\mathbf{s}_i) = 0$, si la variedad en la posición \mathbf{s}_i es CD 101 y $V(\mathbf{s}_i) = 1$, si la variedad en la posición \mathbf{s}_i es CD 103. Fue utilizando el modelo espacial lineal definido en la Ecuaciones (1) y (2). De esta forma, si $Z(\mathbf{s}_i)$ es la productividad de trigo en la posición \mathbf{s}_i , ella es definida como $Z(\mathbf{s}_i) = \beta_0 + \beta_1 V(\mathbf{s}_i) + \varepsilon(\mathbf{s}_i)$, donde β_0 y β_1 son parámetros desconocidos que serán estimados, $V(\mathbf{s}_i)$ es la variedad de trigo en el punto \mathbf{s}_i y $\varepsilon(\mathbf{s}_i)$ es el error aleatorio con $E[\varepsilon(\mathbf{s}_i)] = 0$, y la variación entre puntos en el espacio tiene una estructura de correlación espacial $\Sigma = \varphi_1 I_n + \varphi_2 R(\varphi_3)$, que depende del modelo ajustado a la función de covarianza.

6.2.2 Análisis exploratorios de los datos

En la Tabla 12 son presentadas las estadísticas descriptivas para la variable Productividad de trigo total (PROD) y por variedad (los datos son presentados en el Anexo B). La productividad media del trigo (t ha^{-1}) de la variedad CD 101 fue superior a la variedad CD 103 en $0,208 \text{ t ha}^{-1}$. Los datos son considerados homogéneos en relación a sus respectivas medias en las dos variedades estudiadas (CV menor que 30%). Los coeficientes de asimetría y curtosis, de la distribución de los datos total y por variedad presentaron características de la distribución normal, según los intervalos de 95% de confianza presentados en JONES (1969).

Tabla 12. Estadísticas descriptivas de la Productividad de trigo (PROD) (t ha^{-1}) total y por variedades CD 101 y CD 103

| Estadísticas | PROD-Total | PROD-CD101 | PROD-CD103 |
|--------------|------------|------------|------------|
| n | 83 | 17 | 66 |
| Mín. | 1.480 | 1.880 | 1.480 |
| Q1 | 3.025 | 3.110 | 3.012 |
| Mediana | 3.370 | 3.490 | 3.315 |
| Média | 3.369 | 3.535 | 3.327 |
| Q3 | 3.700 | 4.05 | 3.658 |
| Máx. | 5.950 | 5.950 | 5.280 |
| DP | 0.787 | 1.022 | 0.718 |
| Var | 0.619 | 1.044 | 0.515 |
| CV | 23.36 | 28.92 | 21.576 |
| Ass | 0.249 | 0.471 | -0.105 |
| Curt | 0.876 | -0.151 | 0.502 |

n: número de datos; Mín.: valor mínimo; Máx.: valor máximo; Q1: primer cuartil; Med: mediana; Q3: tercer cuartil; DP: desviación estándar; Var: Varianza; CV: coeficiente de variación; Ass: Coeficiente de asimetría; Curt: Coeficiente de curtosis.

En la Figura 19a es presentado el gráfico *boxplot* que indica la existencia de nueve puntos discrepantes que son 06, 36, 41, 42, 45, 52, 54, 58 y 78 con los valores de productividad de trigo de 5.95, 1.90, 4.85, 1.88, 1.76, 5.28, 1.48, 4.83 y 1.78, $t\ ha^{-1}$, respectivamente, ubicados en el gráfico *postplot* de la Figura 19b.

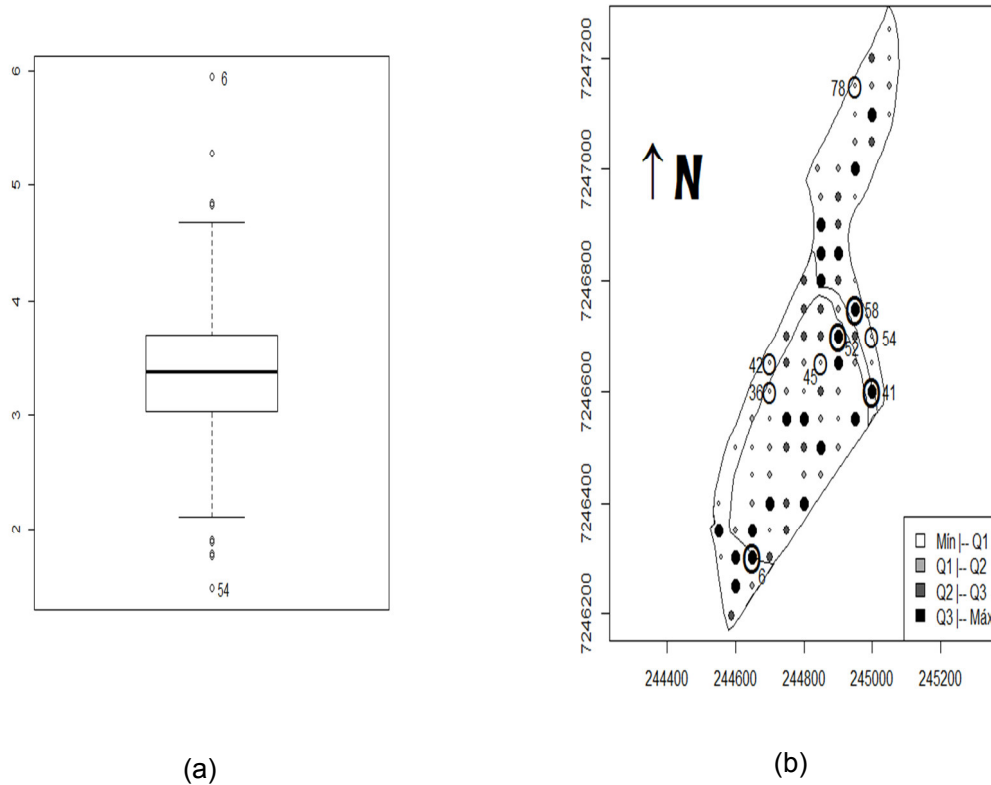


Figura 19. Gráfico *boxplot* (a); gráfico *postplot* (b) para los datos de la Productividad de trigo.

6.2.3 Análisis espacial

Llevando en consideración que la distancia mínima entre puntos de la muestra en la área en estudio fue de 42 m y la máxima fue de 1151 m, en la Figura 20a se presentan los gráficos de las funciones semivarianza muestral en las direcciones de 0°, 45°, 90° y 135°, verificando que los datos son isotrópicos hasta una distancia de 400 m (*cutoff*). La continuidad espacial es verificada en la Figura 20b por medio del semivariograma omnidimensional muestral.

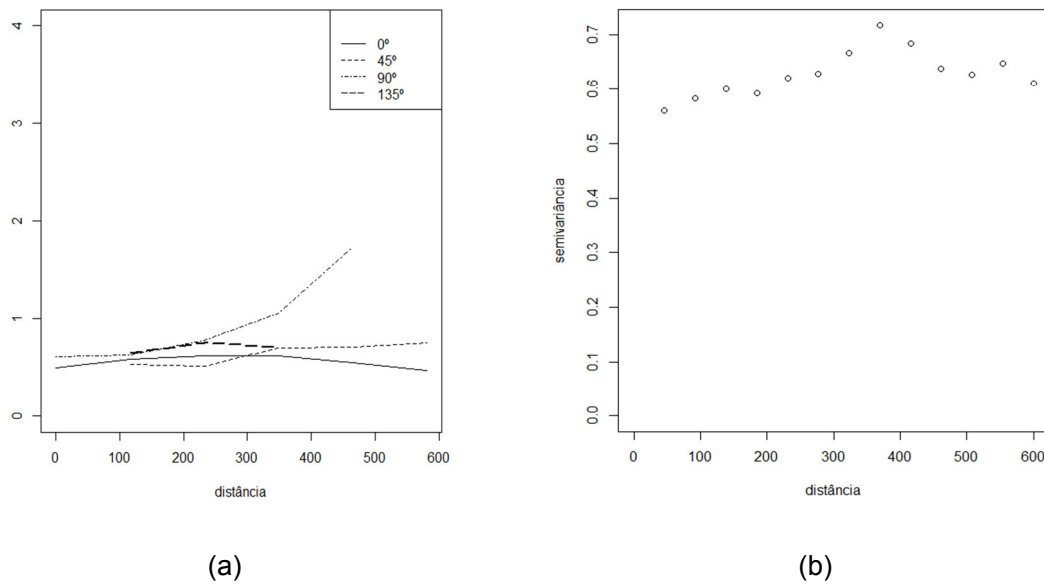


Figura 20. Gráficos de las funciones semivarianza muestral en las direcciones de 0°, 45°, 90° y 135° (a); semivariograma omnidimensional muestral (b).

En la Tabla 13 son presentados los resultados del estudio de la estructura de dependencia espacial de la productividad de trigo con una covariable. En ella es presentada los modelos ajustados: exponencial, gaussiano y familia Matérn con parámetro de orden κ fijo, $\kappa = 0.7, 1.0$ y 1.5 , acompañados de las estimaciones de los parámetros por ML y sus respectivos desvíos estándares (entre paréntesis). Se observa que en todos los modelos ajustados las estimaciones del efecto pepita $\hat{\phi}_1$ son iguales a cero y las estimaciones que definen la contribución (sill) $\hat{\phi}_2$ son semejantes, variando las estimaciones del radio de dependencia espacial (\hat{a}) de 44.34 a 52.16 m. El máximo valor del logaritmo de la función de verosimilitud corresponde al modelo gaussiano.

En la Tabla 14 se presentan los criterios de validación cruzada, AIC y SIC para seleccionar el mejor modelo ajustado a la estructura de variabilidad espacial de la Productividad de trigo. Según los criterios de validación cruzada, AIC e SIC y MLL , el mejor modelo ajustado para la estructura de variabilidad espacial fue el gaussiano. De esta forma el modelo espacial lineal estimado para la productividad de trigo es de la forma: $\hat{Z}(s_i) = 3.4758 - 0.1588V(s_i)$ y los

parámetros estimados de la estructura de variabilidad espacial son: $\hat{\phi}_1 = 0.0$, $\hat{\phi}_2 = 0.6085$, $\hat{\phi}_3 = 32.4594$, con un radio de dependencia espacial de 56.18 m.

Tabla 13. Modelos ajustados por ML, estimativas de los parámetros y desviaciones estándares (entre paréntesis) para el estudio de la variabilidad espacial de la Productividad de trigo

| Modelo | $\hat{\beta}_0$ | $\hat{\beta}_1$ | $\hat{\phi}_1$ | $\hat{\phi}_2$ | $\hat{\phi}_3$ | \hat{a} | <i>MLL</i> |
|--------------|---------------------------------|----------------------------------|---------------------------------|---------------------------------|-----------------------------------|---------------|---------------|
| Exponencial | 3.4928 (0.196) | -0.1764 (0.217) | 0.0000 (1.297) | 0.6054 (1.306) | 18.0968 (235.17) | 54.213 | -96.69 |
| Gausiano | 3.4758 (0.197) | -0.1588 (0.218) | 0.0000 (0.969) | 0.6085 (0.980) | 32.4594 (195.62) | 56.181 | -96.60 |
| Matérn k=0.7 | 3.4909 (0.199) | -0.1744 (0.219) | 0.0000 (0.700) | 0.6056 (0.718) | 15.8771 (158.06) | 54.726 | -96.68 |
| Matérn k=1.0 | 3.4889 (0.203) | -0.1724 (0.221) | 0.0000 (0.377) | 0.6058 (0.412) | 13.7898 (105.79) | 55.139 | -96.66 |
| Matérn k=1.5 | 3.4868 (0.207) | -0.1701 (0.219) | 0.0000 (0.269) | 0.6061 (0.324) | 11.6952 (71.88) | 55.480 | -96.65 |

$\hat{\beta}_0$: estimación del parámetro lineal β_0 ; $\hat{\beta}_1$: estimación del parámetro angular β_1 ; $\hat{\phi}_1$: efecto pepita estimado; $\hat{\phi}_2$: efecto de la contribución estimada; $\hat{\phi}_3$: función del alcance estimada; \hat{a} : alcance estimado (m); *MLL*: máximo valor del logaritmo de la función de verosimilitud, entre paréntesis son presentados las desviaciones estándares.

Tabla 14. Criterios de validación cruzada, *AIC* y *SIC* para la selección del mejor modelo ajustado para la Productividad de trigo

| Modelo | <i>EM</i> | \overline{ER} | <i>S_{EM}</i> | <i>S_{ER}</i> | <i>EA</i> | <i>AIC</i> | <i>SIC</i> |
|-----------------|------------------|-----------------|-----------------------|-----------------------|----------------|-----------------|-----------------|
| Exponencial | -0.001346 | -0.00086 | 0.799612 | 1.019578 | 48.48889 | 203.3726 | -0.00135 |
| Gausiano | -0.002231 | -0.00143 | 0.79617 | 1.019957 | 48.3555 | 203.2014 | -0.00223 |
| Matérn k=0.7 | -0.001423 | -0.0009 | 0.79931 | 1.019603 | 48.48385 | 203.3515 | -0.00142 |
| Matérn k=1.0 | -0.001508 | -0.00096 | 0.798977 | 1.019631 | 48.47626 | 203.3299 | -0.00151 |
| Matérn k=1.5 | -0.001605 | -0.00102 | 0.798588 | 1.019666 | 48.46509 | 203.307 | -0.00161 |

EM: error medio; \overline{ER} : error medio reducido; *S_{EM}*: desviación estándar del error medio; *S_{ER}*: desviación estándar del error medio reducido; *EA*: error absoluto; *AIC*: criterio de información de Akaike; *SIC*: criterio de información de Schwarz.

En la Figura 21 es presentado el mapa temático de la Productividad de trigo ($t\ ha^{-1}$) construido por la interpolación por kriging universal. Se verifican diferencias en la área en estudio según la variedad CD 101 y CD 103 (área 1 variedades CD 101; área 2 y 3 variedad CD 103).

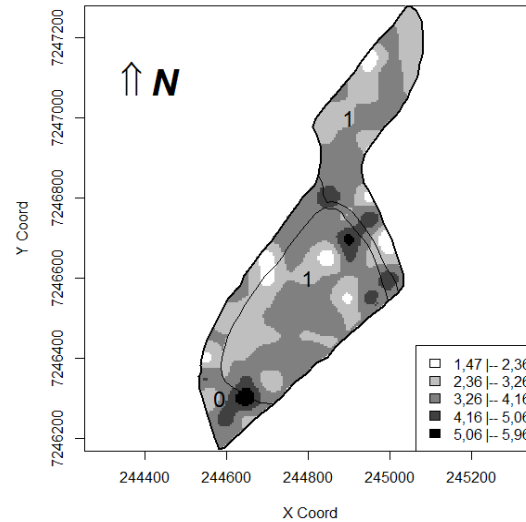
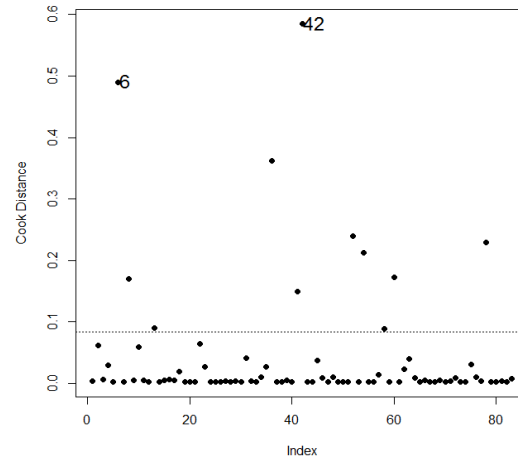


Figura 21. Mapa temático de la Productividad de trigo por áreas, considerando la variedad CD 101 en la subárea 1 y la variedad CD 103 en las subáreas 2 y 3.

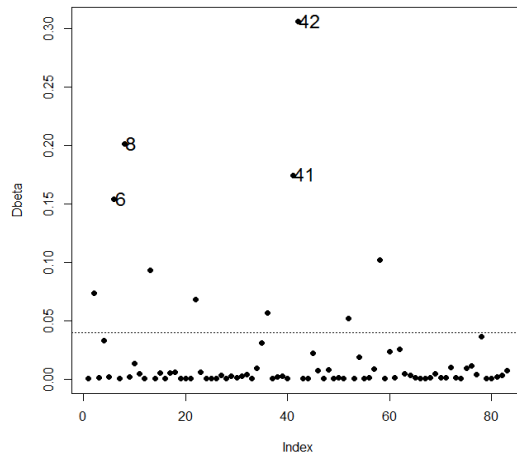
6.2.4 Diagnósticos de Influencia global y local

La distancia de Cook ($D_{i\theta}$) fue aplicada como técnica gráfica de diagnóstico de influencia global y las medidas curvatura normal conformal B_i y $|L_{max}|$ como técnicas gráficas de diagnósticos de influencia local para el estudio de puntos influyentes, utilizado el modelo espacial lineal. Se desea verificar la existencia de observaciones que puedan ejercer influencia en el desplazamiento de la verosimilitud cuando se tiene la perturbación aditiva, perturbación de Zhu y la perturbación de Zhu generalizaran la variable respuesta.

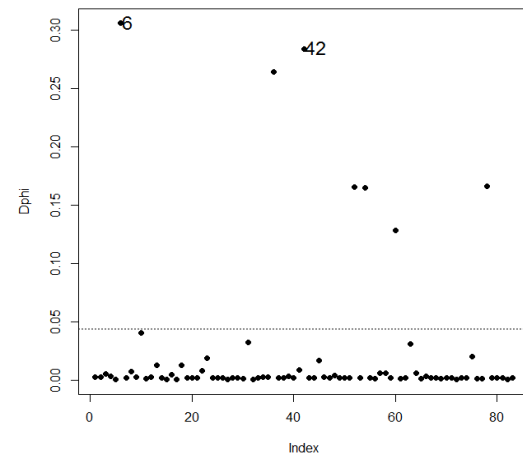
En la Figura 22 presenta el estudio de influencia global utilizando a distancia de Cook $D_{i\theta}$, se observa que los puntos 06 y 42 son considerados como influyentes en los tres gráficos.



(a)



(b)



(c)

Figura 22. Gráfico $D_{i\theta}$ versus orden de los datos (a); Gráfico $D_{i\beta}$ versus orden de los datos (b); Gráfico $D_{i\varphi}$ versus orden de los datos (c) para el estudio de influencia global.

En la Figura 23 se presenta el estudio de influencia local utilizando los gráficos B_i y $|L_{max}|$, como técnica gráfica de influencia local, utilizando el modelo espacial lineal gaussiano ajustado. Se verifica la existencia de las observaciones 08 y 57 como puntos influyentes en el desplazamiento de la verosimilitud cuando se utiliza la perturbación aditiva y la perturbación de Zhu, y las observaciones 09 y 57 como puntos influyentes cuando se utiliza la perturbación de Zhu generalizada en la variable respuesta. Es importante resaltar que ningún de los puntos considerados influyentes fueron identificados como discrepantes, cuando fue analizado el gráfico *boxplot* de la Figura 19a.

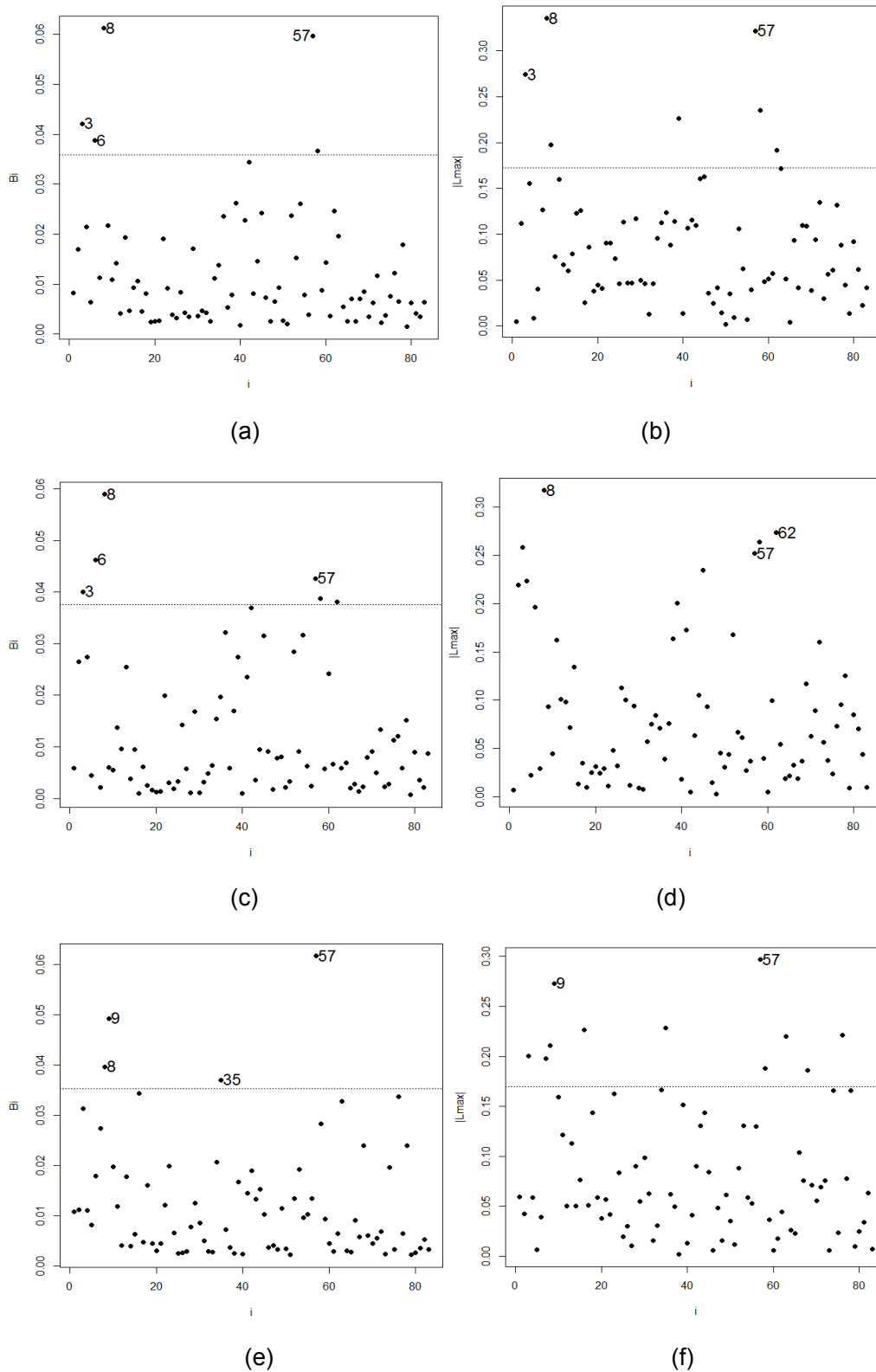


Figura 23. Gráficos de influencia local B_i versus orden (a) y $|L_{max}|$ versus orden (b) para la perturbación aditiva; Gráficos de influencia local B_i versus orden (c) y $|L_{max}|$ versus orden (d) para la perturbación de Zhu; Gráficos de influencia local B_i versus orden (e) y $|L_{max}|$ versus orden (f) para la perturbación de Zhu generalizada.

En la Tabla 15 se presenta un resumen del estudio de la variabilidad espacial sin considerar los puntos influyentes 08 y 57 obtenidos por los esquemas de las perturbaciones aditiva y Zhu y también de presenta un resumen del estudio de la variabilidad espacial sin considerar los puntos influyentes 09 y 57 obtenidos por el esquema de perturbación de Zhu generalizada. La selección del mejor modelo ajustado en cada estudio fue realizado utilizando los criterios de validación cruzada y los índices *AIC* y *SIC*. Se observa que al retirar los puntos influyentes 09 y 57 el nuevo modelo ajustado es el exponencial, alterando las estimaciones de los parámetros por ML y aumentando el radio de dependencia espacial (\hat{a}), en relación al modelo con todos los puntos (modelo gaussiano).

Tabla 15. Modelos ajustados por ML y estimaciones de los parámetros del modelo espacial lineal con sus desviaciones estándares (entre paréntesis), para la Productividad de trigo

| Modelo | $\hat{\beta}_0$ | $\hat{\beta}_1$ | $\hat{\phi}_1$ | $\hat{\phi}_2$ | $\hat{\phi}_3$ | \hat{a} | <i>MLL</i> |
|------------------|-----------------|-----------------|----------------|----------------|----------------|-----------|------------|
| Gausiano-8 | 3.3512 | -0.0438 | 0.00 | 0.6223 | 37.9808 | 65.738 | -94.90 |
| | (0.2060) | (0.2240) | (0.2740) | (0.319) | (149.18) | | |
| Exponencial-57 | 3.5107 | -0.2015 | 0.00 | 0.6136 | 20.4547 | 61.277 | -95.85 |
| | (0.2067) | (0.2278) | (0.7722) | (0.7898) | (245.48) | | |
| Exponencial-9 | 3.4891 | -0.1706 | 0.00 | 0.6129 | 18.4337 | 55.232 | -96,00 |
| | (0.1978) | (0.2203) | (1.2154) | (1.2262) | (238.61) | | |
| Gausiano-8-57 | 3.3846 | -0.0944 | 0.00 | 0.6411 | 40.9917 | 70.949 | -93.22 |
| | (0.2193) | (0.2367) | (0.2444) | (0.3000) | (148.75) | | |
| Exponencial-9-57 | 3.5067 | -0.1949 | 0.00 | 0.6213 | 20.7747 | 62.236 | -95.15 |
| | (0.2087) | (0.2309) | (0.7446) | (0.7637) | (250.45) | | |

$\hat{\beta}_0$: estimación del parámetro lineal β_0 del MEL; $\hat{\beta}_1$: estimación del parámetro angular β_1 del MEL; $\hat{\phi}_1$: efecto pepita estimado; $\hat{\phi}_2$: contribución estimada; $\hat{\phi}_3$: función del alcance estimada; \hat{a} : alcance estimado (m); *MLL*: máximo valor del logaritmo de la función de verosimilitud, entre paréntesis son presentados los desviaciones estándares.

6.2.5 Comparación de mapas temáticos

En la Figura 24 son presentados los mapas temáticos de la productividad de trigo utilizando el modelo espacial lineal y la interpolación por kriging universal. Los mapas temáticos fueron construidos considerando todos los puntos, considerando la eliminación de puntos influyentes 08, 09, 57 y la eliminación de puntos conjuntamente influyentes 08, 57 y 09, 57. Se puede observar algunas diferencias en la construcción de los mapas temáticos. Estas diferencias son analizadas en la Tabla 15 en la construcción de la matriz de errores para el cálculo de los índices de exactitudes (Exactitud Global e índice Kappa).

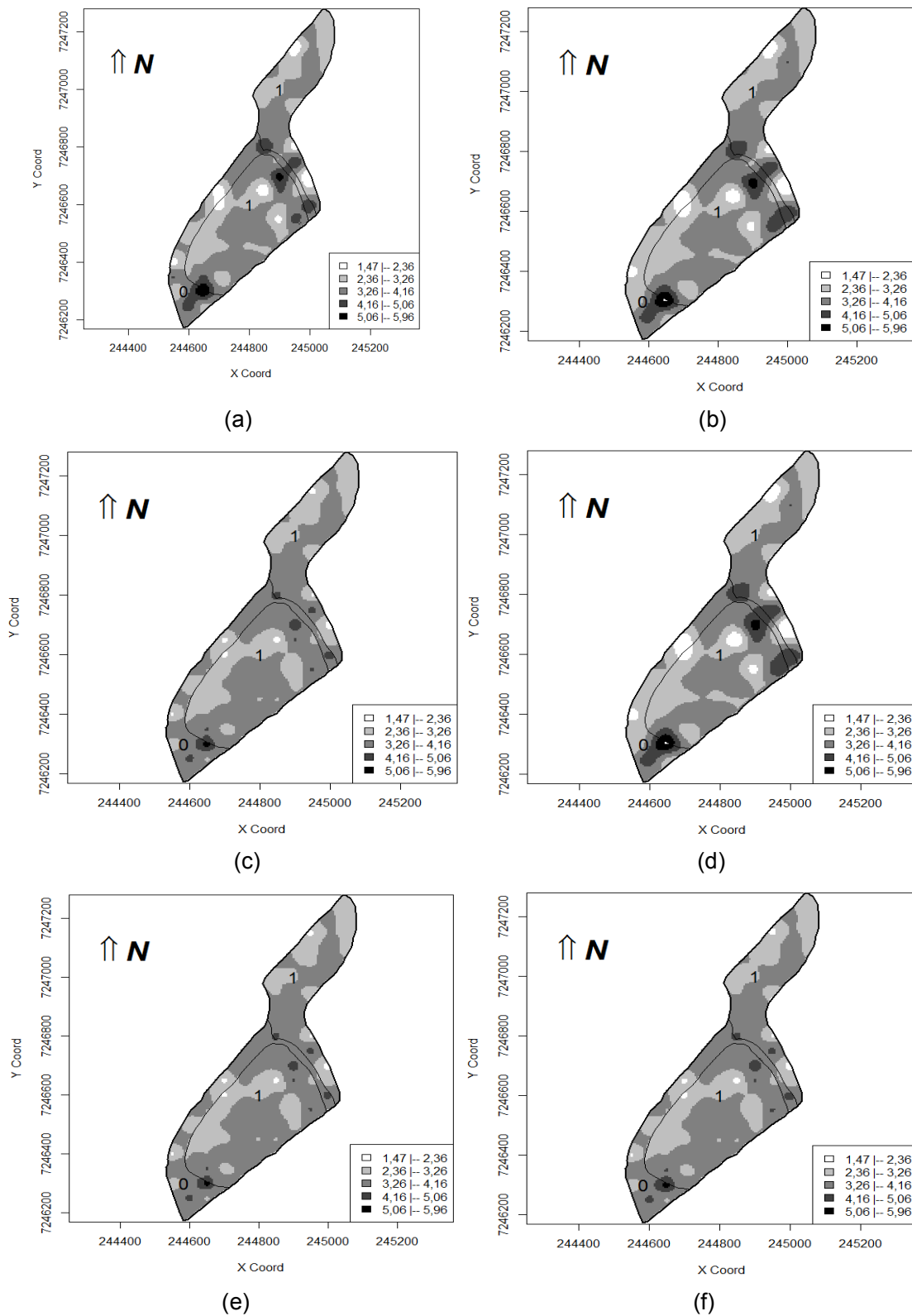


Figura 24. Mapa de la Productividad de trigo por áreas con todos los datos (a); Mapa Productividad de trigo sin la observación 08 (b); Mapa de Productividad de trigo sin la observación 57 (c); Mapa de Productividad de trigo sin la observación 08 y 57 (d); Mapa da Productividad de trigo sin la observación 09 (e) y Mapa de Productividad de trigo sin las observaciones 09 y 57 (f) consideradas influyentes.

En la Tabla 16, se presentan los índices de Exactitud Global (EG) y Kappa (K) calculados a partir de la matriz de errores de cada comparación de mapa construido con todos los puntos (referencia) y con el mapa construido sin los puntos considerados influyentes (modelo) para cada caso. En la comparación de mapa de la productividad de trigo con todos los puntos y del mapa de productividad sin el punto 08 considerado influyente (Gausiano-8), presentaron los índices $EG = 0.9192$ y $K = 0.8689$ las cuales nos indican que no hubo cambio al ser retirado el punto 08 considerado influyente, pues los mapas son considerados semejantes según la clasificación de ANDERSON et al. (2001) y KRIPENDORFF (2004) respectivamente, esto se puede observar en las Figuras 24a y 24b. Lo mismo sucede con la comparación de los mapas al ser retirados conjuntamente los puntos 08 y 57 considerados conjuntamente influyentes por la perturbación aditiva y Zhu, siendo $EG = 0.946$ y $K = 0.8274$. Al ser retirados los puntos considerados influyentes por la perturbación de Zhu generalizada: 09 (Exponencial-9), 57 (Exponencial-57) y ambos (Exponencial-9-57) se verifican en los tres mapas construidos sin estos puntos influyentes (Figuras 24a, 24d, 24c y 24f) dejaron de ser semejantes al mapa construido con todos los puntos (referencia), pues, los índices de exactitud fueron $EG < 0.85$ y $K < 0.80$. Se puede concluir que los puntos 09 y 57 ejercen influencia conjunta en la estructura de variabilidad espacial y en la construcción de los mapas temáticos, pudiendo cambiar el comportamiento de la productividad de trigo en la área en estudio.

Tabla 16. Índices de acurácia: exactitud global (EG) y kappa (K) para la comparación de mapas temáticos de la Productividad de trigo con y sin los puntos influyentes

| Mapa sin los puntos Influyentes/Mapa con todos los puntos | EG | kappa (K) |
|-----------------------------------------------------------|--------|---------------|
| Gausiano-08 | 0.9192 | 0.8663 |
| Exponencial-57 | 0.8427 | 0.7112 |
| Exponencial-09 | 0.8275 | 0.6789 |
| Gausiano-08-57 | 0.8946 | 0.8274 |
| Exponencial-09-57 | 0.8444 | 0.7144 |

Los modelos espaciales lineales permitieron verificar la variabilidad espacial de la productividad de trigo en $t\ ha^{-1}$ en la área en estudio. Se identificaron regiones donde se tienen alta y baja productividad del trigo, que pueden servir para un análisis más adecuado

del suelo por región. Se observa que al retirar las observaciones 09 y 57 consideradas conjuntamente influyentes por la perturbación de Zhu generalizada, provocó cambios en la estimación de los parámetros que definen la estructura de dependencia espacial y en la construcción del mapa temático.

CAPÍTULO 7

CONCLUSIONES

En esta tesis se estudio los métodos geoestadísticos en el estudio de la variabilidad espacial de variables regionalizadas utilizando los modelos espaciales lineales gaussianos.

Los criterios de validación de modelos que fueron presentados en este trabajo fueron de grande importancia como criterios para la selección del mejor modelo ajustado.

El estudio de diagnóstico de influencia global y local, fueron de gran interés para evaluar la influencia de observaciones en los resultados tales como la selección del modelo, las estimación de parámetros y en la construcción de los mapas temáticos.

La aplicación de las técnicas de diagnóstico debe ser parte de todo análisis geoestadística. De esta manera, la información presentada en los mapas temáticos tendrá mejor estimación y mostrará con claridad la estructura de dependencia espacial de los datos.

Las perspectivas para trabajos futuros son de iniciar la construcción de paquetes geoestadísticos a partir de los resultados teóricos obtenidos en este trabajo. Se pretende también dar continuidad al estudio de dependencia espacial tomando r repeticiones en la misma área em estudio de las mismas variables regionalizadas.

BIBLIOGRAFIA

- Anderson, J.F.; Hardy, E. E.; Roach, J.T.; Witmer, R.E. (2001) *A land use and land cover classification system for use with remote sensor data*, U. S. Geological Survey Professional Paper 964, Washington, DC.
- Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. *Journal of the Royal Statistical Society*, London, 51, 469-483.
- Assumpção, R.A.B.; Uribe-Opazo, M.A.; Galea, M. (2011) Local Influence for spatial analysis of soil physical properties and soybean yield Student's t-distribution. *Revista Brasileira de Ciência do Solo*, Viçosa, v.35, n.5, p.1917-1926.
- Bach, M.; Breuer, L.; Frede, H.G.; Huisman, J.A.; Otte, A.; Waldhardt, R. (2006) Accuracy and congruency of three different digital land-use maps. *Landscape and urban planning*. Amsterdam, v.78, n.4, p. 289-299.
- Barnes, R.J. (1991) The variogram sill and the sample variance. *Mathematical geology*, 23, 673-697.
- Borssoi, J.A.; Uribe-Opazo, M.A.; Galea M. (2009) Diagnostic techniques applied in geostatistics for agricultural data analysis. *Revista Brasileira de Ciência do Solo*. Viçosa, v.33, n.6, 1561-1570.
- Burgess, T.M.; Webster, R. (1980) Optimal interpolation and isarithmic mapping of soil properties: I. The semivariogram and punctual kriging. *Journal of Soil Science*, Oxford, v. 31, n. 2, p. 315-331.
- Byrd, R.H.; Lu, P.; Nocedal, J.; Zhu, C. (1995) A limited memory algorithm for bound constrained optimization, *SIAM J. Scientific Computing*, 16, 1190-1208.
- Câmara, G.; Medeiros, J.S. (1998) GIS para meio ambiente. *Sistemas de Informações Geográficas*, Curitiba: Sagres, 140-180.
- Cambardella, C. A.; Moorman, T.B.; Novak, I. M.; Parkin, T.B.; Karlen, D.L.; Turco, R.F.; Konopka, A. E. (1994) Field-scale variability of soil properties in Central Iowa Soils. *Soil Science Society of America Journal*, Madison, v. 58, n. 2, p. 1501-1511.
- Canarache, A. (1994) PENETR – a generalized semi-empirical model estimating soil resistance to penetration. *Soil Till. Res.*, Amsterdam, v.16, n.1, p. 51-70.
- Christensen, R.; Johnson, W.; Pearson, L. (1993) Covariance function diagnostics for spatial linear models, *Mathematical Geology*, 25, 145-160.
- Clark, I. (1979) *Practical Geostatistics*. London, Applied Science Publishers LTD.
- Congalton, R.G.; Green, K. (1999) *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. ISBN 0-87371-986-7, CRC Press.
- Cook, R.D. (1977) Detection of influential observations in linear regression, *Technometrics* 19, 15–18.
- Cook, R.D. (1986) Assessment of local influence, *Journal of the Royal Statistical Society*, B, 48(2), 133-169.

- Cox, D.R.; Hinkley, D.V. (1974). *Theoretical Statistics*. London: Chapman and Hall.
- Cressie, N. (1985) Fitting variogram models by weighted least squares. *Mathematical Geology*, New York, v. 17, n. 4, p. 563-586.
- Cressie, N. (1993) *Statistics for Spatial Data*, John Wiley & Sons.
- Cressie, N.; Hawkins, D.M. (1980) Robust Estimation of the Variogram. *Journal of the international Associate Mathematical Geology*.12, p. 115-125.
- De Bastiani, F.; Uribe-Opazo, M.A.; Dalposso, G.H. (2012) Comparison of maps of spatial variability of soil resistance to penetration constructed with and without covariables using a spatial linear model. *Engenharia Agrícola*, Jaboticabal, v. 32, n.2, p. 394-404.
- De Bastiani, F.; Cysneiros, A.H.M; Uribe-Opazo, M.A.; Galea, M. (2015) Influence diagnostics in elliptical spatial linear models. *Test*, v.24, n.2, p. 322-340.
- Diggle, P. J.; Ribeiro Jr, P. J. (2007) *Model-based geostatistics*, Hardcover.
- Eastman, J.R. (2001) IDRISI 32 Release 2 - Guide To GIS and Image Processing, 2001 – v. 32.20, Clark Labs, Clark University, Worcester, EUA.
- EMBRAPA (2009) Centro Nacional de Pesquisa de Solos (Rio de Janeiro - RJ). *Sistema brasileiro de classificação de solos*. Rio de Janeiro: EMBRAPA – SPI.
- Faraco, M.A.; Uribe-Opazo, M.A.; Silva, E. A. A.; Johann, J.A.; Borssoi, J.A. (2008) Seleção de modelos de variabilidade espacial para elaboração de mapas temáticos de atributos físicos do solo e produtividade da soja, *Revista Brasileira de Ciência do Solo*, 32 (2), 463-476.
- Galea, M.; Paula, G.; Uribe-Opazo, M.A. (2003) On influence diagnostics in univariate elliptical linear regression models, *Statistical Papers*, 44, 23-45.
- Galea, M.; Bolfarine, H.; Vilca, F. (2005) Local influence in comparative calibration models under elliptical t-distributions. *Biometrical Journal*, 47, 691-706.
- Genton, M.G. (1998) Highly robust variogram estimation. *Mathematical Geology*. 30, p. 213-221.
- Gomes, P.F. (2009) *Curso de estatística experimental*. 15 ed. Piracicaba: Degaspari.
- Gradshteyn, I.; Ryzhik, I. (2000) *Tables of Integrals, Series and Products*. Academic Press, New York.
- Guedes, L.P.C.; Uribe-Opazo, M.A.; Johann, J.A.; Souza, E.G. (2008) Anisotropia no estudo da variabilidade espacial de algumas variáveis químicas do solo. *Revista Brasileira de Ciência do Solo*, Viçosa, v. 32, n. 6, p. 2217-2226.
- Isaaks, E.; Srivastava, R. (1989) *An Introduction to Applied Geostatistics*, Oxford University Press, New York.
- Jenness, J.; Wynne, J.J. (2005) Cohen's Kappa and classification table metrics 2.0: an ArcView 3x extension for accuracy assessment of spatially explicit models: U.S. Geological Survey Open-File Report OF 2005-1363. U. S. Geological Survey, Southwest Biological Science Center, Flagstaff, AZ.

Jones, T.A. (1969) Skewness and Kurtosis as criteria of normality in observed frequency distributions. *Journal Sedimentary Petrology*, 39 (4), 1622-1627.

Journel, A. G.; Huijbregts, C.H. (1978) *Mining Geostatistics*, Academic Press, New York.

Journel, A.G. (1989) *Fundamentals of geostatistics in five lessons short course in Geology*. American Geophysical Union, Washington.

Kitanidis, P.K.; Vomvoris, E.G. (1983) A geostatistical approach to the inverse problem in groundwater to the modeling (steady state) and onedimensional. *Water Resources Research*. 19: 677-690.

Krajeski, S.A.; Gibbs, B.L. (1993) *A variogram Primer*. Gibb Associates.

Krige, D.G. (1951) A statistical approach to some basic mine valuation problems on the Witwatersrand, *Journal of the Chemical, Metallurgical and Mining Society of South Africa* 52: 119-139.

Krippendorff, K. (2004) *Content analysis: an introduction to its methodology*, Beverly Hills: Sage Publications.

Lange, K.L.; Little, R. J.A.; Taylor, J.M.G. (1989) Robust statistical modeling using the t distribution, *Journal of the American Statistics*, 84, 881-896.

Ma, Z.; Redmond, R.L. (1995) Tau coefficients for accuracy assessment of classification of remote sensing data. *Photogrammetric Engineering and Remote Sensing*. Bethesda, 61 (4), 453-439.

Mardia, K.V.; Marshall, R.J. (1984) Maximum likelihood models for residual covariance in special regression. *Biometrika*, Oxford, 71 (1), 319-332.

Matérn, B. (1986) Spatial Variation. *Lecture notes in Statistics*, 2ed. New York: Springer.

Matheron, G. (1963) Principles Principe's of Geoestatistics. *Economic Geology*, Lacaster, 58, 1246-1266.

Matheron, G. (1970) Random fuctions and their applications geology. N Merriam, D.F. ed. *Geoestatistics*. 79-88.

McBratney, A.B.; Webster, R. (1986) Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. *Journal Soil Science*, 37, 617-639.

Mello J.M.; Batista, J.L.F.; Ribeiro Jr, P.J.; Oliveira, M.S. (2005) Ajuste e seleção de modelos espaciais de semivariograma visando à estimativa volumétrica de *Eucalyptus randis*. *Scientia Florestalis*, Piracicaba, n. 69, 25-37.

Militino, A.F.; Palacios, M.B.; Ugarte, M.D. (2004) Outliers detection in multivariate spatial linear models. *Journal Statistical Planning and Inference*, 136, 125-146.

Mitchell, A.F.S. (1989) The information matrix, skewness tensor and α -connections for the general multivariate elliptic distribution, *Annals of the Institute of Statistical Mathematics*, 41, 289-304.

Osorio, F. (2006) Diagnósticos de influência em modelos mixtos elípticos com efeitos aleatórios. Tese de Doutorado, Departamento de Estatística, Universidade de São Paulo.

Osorio, F.; Paula, G.A.; Galea, M. (2007) Assessment of Local Influence in Elliptical Linear Models with Longitudinal Structure *Computational Statistics & Data Analysis*, 51 (9), 4354-4368.

Pannatier, Y. (1996) *Variowin 2.2: Software for spatial data analysis in 2D*. New York: Springer.

Paula, G.A. (2013) *Modelos de regressão com apoio computacional*, IME-USP.

Poon, W.; Poon, Y.S. (1999) Conformal normal curvature and assessment of local influence. *Journal of the Royal Statistical Society B*, 61, 51-61.

R Development Core Team. (2013) *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, <http://www.Rproject.org>.

Ribeiro Jr., P.J.; Diggle, P.J. (2013) GeoR: a package for geostatistical analysis. *R-news*, 1 (2), 15-18.

Schwarz, G. (1978) Estimating the dimension of a model. *Annals of Statistics*, v.6, n.2, p.461-464.

Schabenberger, O.; Gotway, C.A. (2005) *Statistical Methods for Spatial Data Analysis*, Chapman & Hall, London.

Uribe-Opazo, M.A.; Kavanagh, C. R. (2006) Estimação de Parâmetros no Ajuste de Modelos Teóricos a Semivariogramas Experimentais. *Revista Brasileira de Estatística*, Rio de Janeiro, 67, 97-122.

Uribe-Opazo, M. A.; Borsoi, J. A.; Galea, M. (2012) Influence Diagnostics in Gaussian Spatial Linear Models, *Journal of Applied Statistics*, 39 (3), 615-630.

Verbeke, G.; Molenberghs, G. (2001) *Linear mixed models for longitudinal data*. Springer, New York.

Viera, S.R. (2000) Geoestatística em estudos de variabilidade espacial do solo. Tópicos em Ciências do Solo. *Revista Brasileira de Ciéncia do Solo*, Viçosa, v. 1, p. 1-54.

Wackernagel, H. (2003) *Multivariate Geostatistics: an introduction with applications*. Germany Springer.

Webster, R.; Oliver, M.A. (2007) *Geostatistics for Environmental Scientists*. Willey, second edition.

Wollenhaupt, N.C.; Mulla, D.J.; Crawford, G. (1997) Soil sampling and interpolation techniques for mapping spatial variability of soil properties. The Site-Specific Management for Agricultural Systems. ASA-CSSA-SSSA, 777. S.Segoe Rd, Madison, W153711, USA 19-53.

Zhu X.; Byrd, R. H.; Nocedal, J. (1997) L-BFGS-B: Algorithm 778: L-BFGS-B, FORTRAN routines for large scale bound constrained optimization, *ACM Transactions on Mathematical Software*, 23 (4), 550-560.

Zhu, H.; Ibrahim, J.G.; Lee, S.; Zhang, H. (2007) Perturbation selection and influence measures in local influence analysis, *The Annals of Statistics*, 35 (6), 2565-2588.

Zhu, H.; Lee, S. (2001) Local influence for incomplete-data models. *Journal of Royal Statistical Society, Series B* 63, 111-126.

Zimmerman, Z. (1989) Computationally efficient restricted maximum likelihood estimation of generalized covariance functions, *Math. Geol.* 21. 655–672.

ANEXOS

ANEXO A

Datos de la resistencia del suelo a la penetración-RSP (Mpa) en la capa**0-0.10 m de profundidad**

| Parcela | X | Y | RSP |
|---------|---------|---------|-------|
| 1 | 40628.4 | 37646.2 | 0.671 |
| 2 | 40674.1 | 37586.8 | 2.697 |
| 3 | 40719.8 | 37527.5 | 3.295 |
| 4 | 40765.5 | 37468.1 | 3.552 |
| 5 | 40811.1 | 37408.8 | 3.692 |
| 6 | 40856.8 | 37349.4 | 3.630 |
| 7 | 40902.5 | 37290.1 | 3.130 |
| 8 | 40948.2 | 37230.7 | 3.124 |
| 9 | 40993.8 | 37171.4 | 3.116 |
| 10 | 40888.5 | 37185.3 | 4.195 |
| 11 | 40842.8 | 37244.6 | 3.554 |
| 12 | 40797.1 | 37304.0 | 2.350 |
| 13 | 40751.4 | 37363.3 | 3.125 |
| 14 | 40705.8 | 37422.7 | 2.864 |
| 15 | 40660.1 | 37482.1 | 2.662 |
| 16 | 40568.7 | 37600.7 | 3.511 |
| 17 | 40523.1 | 37660.1 | 2.681 |
| 18 | 40463.4 | 37614.6 | 2.920 |
| 19 | 40509.1 | 37555.3 | 2.680 |
| 20 | 40600.4 | 37436.6 | 1.398 |
| 21 | 40737.4 | 37258.5 | 2.023 |
| 22 | 40783.1 | 37199.2 | 2.016 |
| 23 | 40828.8 | 37139.8 | 2.652 |
| 24 | 40677.7 | 37213.1 | 3.455 |
| 25 | 40586.4 | 37331.8 | 2.976 |
| 26 | 40540.7 | 37391.2 | 2.367 |
| 27 | 40495.0 | 37450.5 | 2.403 |
| 28 | 40449.4 | 37509.9 | 2.560 |
| 29 | 40403.7 | 37569.2 | 3.259 |
| 30 | 40344.0 | 37523.8 | 3.651 |
| 31 | 40389.7 | 37464.4 | 1.889 |
| 32 | 40435.4 | 37405.1 | 2.170 |
| 33 | 40481.0 | 37345.7 | 3.672 |
| 34 | 40526.7 | 37286.4 | 2.876 |
| 35 | 40572.4 | 37227.0 | 3.423 |

| Parcela | X | Y | RSP |
|---------|---------|---------|-------|
| 46 | 40270.3 | 37373.6 | 2.320 |
| 47 | 40316.0 | 37314.2 | 2.556 |
| 48 | 40361.7 | 37254.9 | 2.469 |
| 49 | 40407.3 | 37195.5 | 2.640 |
| 50 | 40347.7 | 37150.1 | 3.590 |
| 51 | 40302.0 | 37209.4 | 1.981 |
| 52 | 40256.3 | 37268.8 | 2.995 |
| 53 | 40210.6 | 37328.1 | 2.392 |
| 54 | 40164.9 | 37387.5 | 1.698 |
| 55 | 40119.3 | 37446.8 | 0.999 |
| 56 | 40073.6 | 37506.2 | 2.237 |
| 57 | 40027.9 | 37565.5 | 2.374 |
| 58 | 39876.9 | 37638.8 | 2.640 |
| 59 | 39922.5 | 37579.4 | 2.750 |
| 60 | 39968.2 | 37520.1 | 2.080 |
| 61 | 40013.9 | 37460.7 | 2.992 |
| 62 | 40105.3 | 37342.0 | 1.851 |
| 63 | 40150.9 | 37282.7 | 2.418 |
| 64 | 40196.6 | 37223.3 | 2.082 |
| 65 | 40242.3 | 37164.0 | 3.160 |
| 66 | 40288.0 | 37104.6 | 2.813 |
| 67 | 40182.6 | 37118.5 | 2.926 |
| 68 | 40136.9 | 37177.9 | 2.388 |
| 69 | 40091.3 | 37237.2 | 2.692 |
| 70 | 40045.6 | 37296.6 | 2.805 |
| 71 | 39999.9 | 37355.9 | 2.455 |
| 72 | 39908.5 | 37474.7 | 2.369 |
| 73 | 39862.8 | 37534.0 | 1.929 |
| 74 | 39817.2 | 37593.3 | 2.537 |
| 75 | 39757.5 | 37547.9 | 3.004 |
| 76 | 39803.2 | 37488.6 | 3.330 |
| 77 | 39848.8 | 37429.2 | 2.028 |
| 78 | 39894.5 | 37369.9 | 2.763 |
| 79 | 39940.2 | 37310.5 | 2.329 |
| 80 | 39985.9 | 37251.2 | 3.090 |

| | | | |
|----|---------|---------|-------|
| 36 | 40512.7 | 37181.6 | 2.416 |
| 37 | 40467.0 | 37240.9 | 2.519 |
| 38 | 40421.4 | 37300.3 | 1.808 |
| 39 | 40375.7 | 37359.6 | 2.700 |
| 40 | 40330.0 | 37419.0 | 2.966 |
| 41 | 40284.3 | 37478.3 | 3.508 |
| 42 | 40238.6 | 37537.7 | 3.481 |
| 43 | 40133.3 | 37551.6 | 3.362 |
| 44 | 40179.0 | 37492.3 | 2.102 |
| 45 | 40224.6 | 37432.9 | 2.389 |

| | | | |
|----|---------|---------|-------|
| 81 | 40031.6 | 37191.8 | 2.741 |
| 82 | 40077.3 | 37132.5 | 2.544 |
| 83 | 40122.9 | 37073.1 | 3.245 |
| 84 | 40017.6 | 37087.0 | 3.041 |
| 85 | 39971.9 | 37146.4 | 4.064 |
| 86 | 39926.2 | 37205.7 | 1.304 |
| 87 | 39880.5 | 37265.1 | 2.913 |
| 88 | 39834.8 | 37324.4 | 2.541 |
| 89 | 39789.2 | 37383.8 | 4.269 |
| 90 | 39743.5 | 37443.1 | 4.269 |
| 91 | 39697.8 | 37502.5 | 2.710 |
| 92 | 39683.8 | 37397.7 | 2.393 |
| 93 | 39729.5 | 37338.3 | 2.627 |

Borde de la área en estudio.

| X | Y |
|----------|----------|
| 40610.35 | 37713.08 |
| 40582.67 | 37713.08 |
| 40568.83 | 37708.07 |
| 40544.61 | 37700.54 |
| 40532.50 | 37698.04 |
| 40513.48 | 37698.04 |
| 40496.18 | 37693.02 |
| 40478.88 | 37682.99 |
| 40463.31 | 37677.98 |
| 40451.20 | 37667.95 |
| 40435.64 | 37657.92 |
| 40416.61 | 37647.89 |
| 40401.04 | 37635.35 |
| 40375.09 | 37620.31 |
| 40357.80 | 37607.77 |
| 40340.50 | 37600.25 |
| 40328.39 | 37592.72 |
| 40312.82 | 37590.22 |
| 40300.71 | 37580.19 |
| 40283.41 | 37575.17 |
| 40267.85 | 37570.16 |
| 40238.44 | 37567.65 |
| 40224.60 | 37562.63 |
| 40184.82 | 37557.62 |
| 40164.06 | 37557.62 |

ANEXO B

Datos de la productividad de trigo ($t\ ha^{-1}$) por variedad

| Parcela | X | Y | Prod-Trigo | Variedad |
|---------|--------|---------|------------|----------|
| 1 | 244588 | 7246195 | 3.40 | 0 |
| 2 | 244600 | 7246250 | 4.67 | 0 |
| 3 | 244650 | 7246250 | 3.11 | 0 |
| 4 | 244558 | 7246300 | 2.42 | 0 |
| 5 | 244600 | 7246300 | 4.05 | 0 |
| 6 | 244650 | 7246300 | 5.95 | 0 |
| 7 | 244700 | 7246300 | 3.61 | 1 |
| 8 | 244550 | 7246350 | 4.18 | 0 |
| 9 | 244600 | 7246350 | 3.28 | 1 |
| 10 | 244650 | 7246350 | 4.31 | 1 |
| 11 | 244700 | 7246350 | 2.82 | 1 |
| 12 | 244750 | 7246350 | 3.38 | 1 |
| 13 | 244550 | 7246400 | 2.11 | 0 |
| 14 | 244650 | 7246400 | 3.25 | 1 |
| 15 | 244700 | 7246400 | 3.91 | 1 |
| 16 | 244750 | 7246400 | 3.62 | 1 |
| 17 | 244800 | 7246400 | 4.01 | 1 |
| 18 | 244650 | 7246450 | 2.71 | 1 |
| 19 | 244700 | 7246450 | 3.07 | 1 |
| 20 | 244800 | 7246450 | 3.22 | 1 |
| 21 | 244850 | 7246450 | 3.24 | 1 |
| 22 | 244600 | 7246500 | 2.63 | 0 |
| 23 | 244650 | 7246500 | 2.68 | 1 |
| 24 | 244700 | 7246500 | 3.35 | 1 |
| 25 | 244750 | 7246500 | 3.58 | 1 |
| 26 | 244800 | 7246500 | 3.56 | 1 |
| 27 | 244850 | 7246500 | 3.98 | 1 |
| 28 | 244900 | 7246500 | 3.10 | 1 |
| 29 | 244650 | 7246550 | 3.25 | 0 |
| 30 | 244700 | 7246550 | 3.00 | 1 |
| 31 | 244750 | 7246550 | 3.88 | 1 |
| 32 | 244800 | 7246550 | 4.07 | 1 |
| 33 | 244850 | 7246550 | 3.37 | 1 |
| 34 | 244900 | 7246550 | 2.22 | 1 |
| 35 | 244950 | 7246550 | 4.45 | 1 |
| 36 | 244700 | 7246600 | 1.90 | 1 |
| 37 | 244750 | 7246600 | 3.21 | 1 |
| 38 | 244800 | 7246600 | 2.96 | 1 |
| 39 | 244850 | 7246600 | 3.40 | 1 |

| | | | | |
|----|--------|---------|------|---|
| 40 | 244900 | 7246600 | 3.21 | 1 |
| 41 | 245000 | 7246600 | 4.85 | 0 |
| 42 | 244700 | 7246650 | 1,88 | 0 |
| 43 | 244750 | 7246650 | 3,39 | 1 |
| 44 | 244800 | 7246650 | 3,14 | 1 |
| 45 | 244850 | 7246650 | 1.76 | 1 |
| 46 | 244900 | 7246650 | 4.35 | 1 |
| 47 | 244950 | 7246650 | 3.14 | 1 |
| 48 | 245000 | 7246650 | 2.56 | 1 |
| 49 | 244750 | 7246700 | 3.57 | 0 |
| 50 | 244800 | 7246700 | 3.56 | 1 |
| 51 | 244850 | 7246700 | 3.47 | 1 |
| 52 | 244900 | 7246700 | 5.28 | 1 |
| 53 | 244950 | 7246700 | 3.55 | 0 |
| 54 | 245000 | 7246700 | 1.48 | 1 |
| 55 | 244800 | 7246750 | 3.49 | 0 |
| 56 | 244850 | 7246750 | 3.64 | 1 |
| 57 | 244900 | 7246750 | 3.33 | 0 |
| 58 | 244950 | 7246750 | 4.83 | 1 |
| 59 | 244800 | 7246800 | 3.66 | 0 |
| 60 | 244850 | 7246800 | 4.68 | 1 |
| 61 | 244900 | 7246800 | 3.65 | 1 |
| 62 | 244950 | 7246800 | 2.13 | 1 |
| 63 | 244850 | 7246850 | 3.91 | 1 |
| 64 | 244900 | 7246850 | 3.95 | 1 |
| 65 | 244850 | 7246900 | 3.74 | 1 |
| 66 | 244900 | 7246900 | 3.66 | 1 |
| 67 | 244850 | 7246950 | 3.15 | 1 |
| 68 | 244900 | 7246950 | 3.53 | 1 |
| 69 | 244950 | 7246950 | 2.79 | 1 |
| 70 | 244840 | 7247000 | 3,05 | 1 |
| 71 | 244900 | 7247000 | 3.08 | 1 |
| 72 | 244950 | 7247000 | 4.02 | 1 |
| 73 | 244950 | 7247050 | 3.05 | 1 |

Variedad CD 101 = 0 y Variedad CD 103=1.